



Garcia Pintos, L. P., Linden, N., Malabarba, A., Short, T., & Winter, A. J. (2017). Equilibration Time Scales of Physically Relevant Observables. *Physical Review X*, 7(3), [031027].  
<https://doi.org/10.1103/PhysRevX.7.031027>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1103/PhysRevX.7.031027](https://doi.org/10.1103/PhysRevX.7.031027)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via APS at <https://link.aps.org/doi/10.1103/PhysRevX.7.031027> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Equilibration Time Scales of Physically Relevant Observables

Luis Pedro García-Pintos,<sup>1,2,\*</sup> Noah Linden,<sup>1,†</sup> Artur S. L. Malabarba,<sup>3,‡</sup> Anthony J. Short,<sup>3,§</sup> and Andreas Winter<sup>4,¶</sup>

<sup>1</sup>*School of Mathematics, University of Bristol, University Walk, Bristol BS8 1 TW, United Kingdom*

<sup>2</sup>*Institute for Quantum Studies, Chapman University, 1 University Drive, Orange, California 92866, USA*

<sup>3</sup>*H.H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, BS8 1TL, United Kingdom*

<sup>4</sup>*ICREA & Física Teòrica: Informació i Fenòmens Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain*

(Received 12 December 2015; revised manuscript received 11 May 2017; published 10 August 2017)

We address the problem of understanding, from first principles, the conditions under which a quantum system equilibrates rapidly with respect to a concrete observable. On the one hand, previously known general upper bounds on the time scales of equilibration were unrealistically long, with times scaling linearly with the dimension of the Hilbert space. These bounds proved to be tight since particular constructions of observables scaling in this way were found. On the other hand, the computed equilibration time scales for certain classes of typical measurements, or under the evolution of typical Hamiltonians, are unrealistically short. However, most physically relevant situations fall outside these two classes. In this paper, we provide a new upper bound on the equilibration time scales which, under some physically reasonable conditions, give much more realistic results than previously known. In particular, we apply this result to the paradigmatic case of a system interacting with a thermal bath, where we obtain an upper bound for the equilibration time scale independent of the size of the bath. In this way, we find general conditions that single out observables with realistic equilibration times within a physically relevant setup.

DOI: [10.1103/PhysRevX.7.031027](https://doi.org/10.1103/PhysRevX.7.031027)

Subject Areas: Quantum Physics,  
Quantum Information,  
Statistical Physics

## I. INTRODUCTION

Knowing the details of how systems approach equilibrium is a major topic within statistical mechanics. However, deriving results on the equilibration time scales that are both general and apply to physically relevant situations has proven to be a challenge; this is one of the major open problems in understanding equilibration processes of quantum systems.

This paper addresses the time scales for reaching equilibrium in closed quantum systems. Recently, there have been promising advances [1–5], that add to the vast amount of literature on these issues in more specific models [6–19] (for recent thorough reviews of this and related topics, see Refs. [20–22]). In particular, we have learned that typical observables (when appropriately drawn at random) equilibrate rapidly [2,3,5] and that the same is

true for the evolution under typical Hamiltonians [9–11] and for systems starting from typical nonequilibrium subspaces [4]. Remarkably, this rapid equilibration has even been observed experimentally in certain systems [5]. Yet, one can construct observables that take an extremely long time to approach equilibrium, up to a time proportional to the Hilbert space dimension of the system [2,3]. Note that by fast-vs-slow equilibration we do not mean picoseconds vs years; slow can refer to time scales longer than the “age of the universe” for the constructions found in the papers mentioned above.

It is important to note that the above-mentioned results do not teach us a great deal about what happens for a given physically relevant observable. For instance, they do not tell us what the time scales of equilibration for a system interacting with an environment are. Meanwhile, the typical (in the mathematical sense) measurements considered will generally not represent physically relevant observables.

Moreover, the fact that one can always find mathematical constructions of observables that equilibrate after extremely long times, as in Refs. [2] and [3], implies that extra—more physical—conditions are fundamental in singling out the observables that equilibrate within reasonable time scales.

In this paper, we consider the following physically relevant scenario—measurements on a small system that is interacting with a large, highly mixed bath via a given (nonrandom) Hamiltonian. The main result is to find

\*Corresponding author.

[lpgarciapintos@gmail.com](mailto:lpgarciapintos@gmail.com)

<sup>†</sup>[n.linden@bristol.ac.uk](mailto:n.linden@bristol.ac.uk)

<sup>‡</sup>[artur.malabarba@bristol.ac.uk](mailto:artur.malabarba@bristol.ac.uk)

<sup>§</sup>[tony.short@bristol.ac.uk](mailto:tony.short@bristol.ac.uk)

<sup>¶</sup>[andreas.winter@uab.cat](mailto:andreas.winter@uab.cat)

*Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

sufficient conditions on the initial state, observable, and Hamiltonian that ensure reasonably fast time scales, in particular, time scales that do not grow or decrease with the full dimension of the Hilbert space. We find that this is the case for sufficiently mixed initial states (such as thermal states of not-too-low temperature)—provided some natural conditions on the off-diagonal matrix elements of the observable and initial state in the energy basis are met—that essentially ensure that a wide range of frequencies are involved in the evolution. This will be applied to the paradigmatic case of a small system interacting with a thermal bath in the microcanonical ensemble [23], where we obtain an upper bound on the equilibration time scale which does not depend on the dimension of the bath. Importantly, the results obtained here do not depend on particular details of the system under consideration.

We say that a system equilibrates when it approaches some steady state, and remains close to it, for some reasonably long time interval [24]. Given the fact that for finite-dimensional systems there always exist revivals—times (in general, very long) in which the system returns arbitrarily close to the initial state—in quantum mechanics, one cannot have equilibration in the strict sense. Therefore, following Ref. [25], we say that a system equilibrates if, for most times, its state is close to some fixed steady state. This fixed steady state is then called the equilibrium state.

Here, this closeness is assessed with respect to some particular observable  $A$ , so we say equilibration takes place if  $A$  cannot distinguish the instantaneous state from the equilibrium one. Restricting to different kinds of observables leads to different notions. Then, for instance, an observable acting on a subsystem probes whether that subsystem has equilibrated, and what happens in the remainder of the closed system is only relevant in how it affects the evolution of this subsystem. However, taking other kinds of observables, for example,  $A$  being some many-body observable, gives a different view of the process. These sort of questions are particularly relevant since experiments are bringing mesoscopic quantum systems closer to observation [26–30]. Notice that these situations are, in general, not described by master equations, and usually one needs to solve the actual evolution of the system in order to learn about time scales of equilibration.

We start in Sec. II by introducing the necessary notions for this paper and a statement of the main result. Section III contains a general upper bound on the time-averaged distance between instantaneous and equilibrium states, and an analysis of the time decay of this bound. Section IV contains an expression for the time scale of equilibration, which depends on the observable, state, and Hamiltonian under consideration, the first main proof in the paper. In Sec. V, we apply the result to the case of a system interacting with a thermal bath in the microcanonical ensemble, an important application of the previous part.

We end in Sec. VI with an analysis of the conditions necessary to obtain reasonably fast equilibration. All detailed calculations can be found in the appendixes.

## II. SETTING AND SPECIAL CASES OF THE MAIN RESULTS

Consider a closed quantum system with a Hamiltonian  $H$  and an initial state given by the density matrix  $\rho_0$  in a Hilbert space  $\mathcal{H}$ . We start by focusing on a weak notion of distance between states, based on comparing the instantaneous expected value of an observable  $A$  to its equilibrium expected value,

$$\tilde{D}_A(\rho_t, \omega) = \frac{|\text{Tr}[\rho_t A] - \text{Tr}[\omega A]|^2}{4\|A\|^2}, \quad (1)$$

where the evolved state is  $\rho_t = e^{-iHt}\rho_0 e^{iHt}$ , and  $\omega = \langle \rho_t \rangle_\infty$  is the equilibrium state [25,31], where  $\langle f(t) \rangle_T = (1/T) \int_0^T f(t) dt$  denotes a time average. Note that the equilibrium state is simply the initial state decohered in the energy basis since the infinite time averaging removes any oscillating terms. The operator  $A$  is assumed to be Hermitian, with  $\|A\|$  denoting its spectral norm [32]. With this definition,  $0 \leq \tilde{D}_A \leq 1$ . For simplicity, we take units such that  $\hbar = 1$ .

Obviously, equilibration of expectation values does not imply equilibration in general, since one can have very different distributions with the same expected values. However, even for this weak notion of equilibration, no reasonable time-scale bounds for physically relevant observables were known up to now. Furthermore, it is easy to extend our calculations to a stricter notion of equilibration, the distinguishability between  $\rho_t$  and  $\omega$ , given a measurement of  $A$  (for completeness, we show this in Appendix A). In order to distinguish the quantity  $\tilde{D}_A(\rho_t, \omega)$  from the actual distinguishability, we call it the weak distinguishability.

We can express the time average of the weak distinguishability as

$$\begin{aligned} \langle \tilde{D}_A(\rho_t, \omega) \rangle_T &= \frac{1}{4} \left\langle \left| \sum_{j,k} e^{-i(E_j - E_k)t} (\rho_{jk} - \omega_{jk}) \frac{A_{kj}}{\|A\|} \right|^2 \right\rangle_T \\ &= \frac{1}{4} \left\langle \left| \sum_{\alpha} v_{\alpha} e^{-iG_{\alpha}t} \right|^2 \right\rangle_T \\ &= \frac{1}{4} \sum_{\alpha\beta} v_{\alpha} v_{\beta}^* \langle e^{-i(G_{\alpha} - G_{\beta})t} \rangle_T, \end{aligned} \quad (2)$$

where energy levels are denoted by  $E_j$ , and the matrix elements of the initial state, equilibrium state, and observable in the energy basis are  $\rho_{jk}$ ,  $\omega_{jk}$ , and  $A_{jk}$ , respectively [1]. The index  $\alpha$  represents pairs  $(j, k)$  of levels with

distinct energies; we denote the corresponding energy gap by  $G_\alpha = (E_j - E_k)$  and define the coefficients

$$v_\alpha = v_{(j,k)} = \rho_{jk} \frac{A_{kj}}{\|A\|}. \quad (3)$$

Notice that only terms with nonzero energy gaps appear in the sum in Eq. (2) since  $\omega_{jk} = \rho_{jk}$  for  $E_j = E_k$ .

Our aim is to prove that the time average of the weak distinguishability considered above becomes small. Since  $\tilde{D}$  is a positive quantity, this would allow us to conclude that, for most times, the weak distinguishability is small, showing that equilibration occurs. The main objective of this paper is to determine, or at least to upper bound, the time scale  $T_{\text{eq}}$  in which this decay happens.

The following normalized distribution will be crucial for our proofs:

$$p_\alpha \equiv \frac{|v_\alpha|}{Q} = \frac{1}{Q} \frac{|\rho_{jk} A_{kj}|}{\|A\|}, \quad (4)$$

with the normalization factor

$$Q \equiv \sum_\alpha |v_\alpha| = \sum_{jk: E_j \neq E_k} |\rho_{jk}| \frac{|A_{kj}|}{\|A\|}. \quad (5)$$

The distribution  $p_\alpha$  contains information of all the physical quantities relevant for the dynamics, namely, the observable  $A$ , the initial state  $\rho_0$ , and the Hamiltonian  $H$ , and it is a measure of the significance of the different energy gaps  $G_\alpha$ .

Our main technical result is a general bound on equilibration times for observables when the initial state is highly mixed (Theorem 6). Before embarking on the proofs of our general technical results, it may be illuminating to see how they apply in certain special cases that are of physical interest. The first case concerns a small system interacting with a bath that is in a maximally mixed state. The second case is a version of our main physical theorem (Theorem 8), in which the bath is in a microcanonical state.

Let us first consider a small system  $S$  of dimension  $d_S$  interacting with a large bath in the maximally mixed state  $\rho_B = [(\mathbb{1}_B)/d_B^\Delta]$ . We can then prove the following [this follows straightforwardly from Theorem 6 by taking  $A = A_S \otimes \mathbb{1}_B$  and  $Q$  bounded by Eq. (29)].

**Theorem 1** (Bound for a system interacting with a maximally mixed bath). For any system observable  $A = A_S \otimes \mathbb{1}_B$ , initial state  $\rho_0 = \rho_S \otimes [(\mathbb{1}_B)/d_B^\Delta]$ , and Hamiltonian  $H = H_B + H_S + H_I$ ,

$$\langle \tilde{D}_A(\rho_t, \omega) \rangle_T \leq \frac{\pi a(\epsilon) \|A_S\|^{1/2} Q^{5/2}}{T \sqrt{|\text{Tr}([\rho_0, H_S + H_I], H_S + H_I) A_S|}} + \pi \delta(\epsilon) Q^2, \quad (6)$$

where

$$Q \leq \sqrt{d_S \text{Tr}_S[\rho_S^2]}, \quad (7)$$

and  $a(\epsilon)$  and  $\delta(\epsilon)$  depend on the distribution  $p_\alpha$  and an arbitrary parameter  $\epsilon > 0$ . They are described briefly below and defined in Proposition 5.

Crucially, we show in Sec. III that if the initial state  $\rho_0$ , observable  $A$ , and Hamiltonian  $H$  are such that  $p_\alpha$  is spread over many different energy gaps and approximately unimodal, then we can choose  $\epsilon$  such that  $\delta(\epsilon) \ll 1$  and  $a(\epsilon) \sim 1$ . We argue that this is to be expected for a wide range of initial states in systems with interacting Hamiltonians, and in Appendix F, we show it explicitly in a simulation of a spin ring, i.e., a 1D Ising model with transversal magnetic field and periodic boundary conditions.

Moreover, we will argue in Sec. IV B that we would expect to achieve a reduction in  $\delta(\epsilon)$  as the size of the bath increases; hence, the second term in Eq. (6) becomes small for large baths, and we find that equilibration occurs for large enough times  $T$ .

We can think of Theorem 1 as describing the system coupled to an infinite-temperature bath. To extend the analysis to a more physically realistic finite-temperature bath (with inverse temperature  $\beta$ ), we consider a bath that is initially in the microcanonical ensemble. Hence, the bath starts in a state  $\rho_B = [(\mathbb{1}_B^\Delta)/d_B^\Delta]$ , where  $\mathbb{1}_B^\Delta$  is the projector on some microcanonical window of width  $\Delta$  and dimension  $d_B^\Delta$ .

We can then prove the following. (This is Theorem 8 in Sec. IV, applied to the special case in which  $A = A_S \otimes \mathbb{1}_B$ . Theorem 8 also applies to general observables.)

**Theorem 2** (Bound for a system interacting with a thermal bath). For any system observable  $A = A_S \otimes \mathbb{1}_B$ , initial state  $\rho_0 = \rho_S \otimes [(\mathbb{1}_B^\Delta)/d_B^\Delta]$ , Hamiltonian  $H = H_B + H_S + H_I$ , and any  $K > 0$  and  $\epsilon > 0$ , the weak distinguishability satisfies

$$\langle \tilde{D}_A(\rho_t, \omega) \rangle_T \leq \frac{\pi a(\epsilon) \|A_S\|^{1/2} Q_2^{5/2}}{T \sqrt{|\text{Tr}([\rho_0, H_S + H_I], H_S + H_I) A_S|}} + \pi \delta(\epsilon) Q_2^2 + \frac{18}{K^2}, \quad (8)$$

where  $a(\epsilon)$  and  $\delta(\epsilon)$  are defined in Proposition 5. For a bath with density of states proportional to  $e^{\beta E}$  in the vicinity of the microcanonical window,

$$Q_2 \leq \sqrt{\frac{d_S \text{Tr}_S[\rho_S^2] e^{\beta \|H_S\| + (1 + \sqrt{2d_S}) K \beta \|H_I\|}}{(1 - \frac{1}{K})(1 - e^{-\beta \Delta})}} + \frac{2}{K}. \quad (9)$$

Considering a sufficiently large bath and choosing the constant  $K$  such that the last term is also small, we obtain that equilibration eventually occurs for large enough times  $T$ .

More precisely, we find that the system will be equilibrated with respect to  $A$ , in the sense described above, for times  $T \gg T_{\text{eq}}$ , where

$$T_{\text{eq}} \equiv \frac{\pi a(\epsilon) \|A_S\|^{1/2} Q_2^{5/2}}{\sqrt{|\text{Tr}([\rho_0, H_S + H_I], H_S + H_I) A_S|}}. \quad (10)$$

Crucially, note that if interactions between the system and bath are short-ranged, such that  $H_I$  only couples the system to a finite region in the bath (e.g., nearest-neighbor interactions in a spin lattice), then  $T_{\text{eq}}$  does not scale with the size of the bath. Instead, it depends on details of the system and its coupling to the environment, and it can be easily calculated from the initial state, the observable, and the Hamiltonian once  $a(\epsilon)$  has been estimated.

In the next section, we will show a general bound on the weak distinguishability, setting the groundwork for proving the above results.

### III. GENERAL BOUND ON AVERAGE DISTANCE

Since  $\tilde{D}_A$  is a positive quantity, it satisfies

$$\langle \tilde{D}_A(\rho_t, \omega) \rangle_T \leq \frac{5\pi}{4} \langle \tilde{D}_A(\rho_t, \omega) \rangle_{L_T}, \quad (11)$$

where  $\langle f(t) \rangle_{L_T} \equiv \int_{-\infty}^{\infty} \{ [f(t)T] / [T^2 + (t-T/2)^2] \} \frac{dt}{\pi}$  denotes the Lorentzian time average of the function  $f$  [3]. Upper bounding the sum in Eq. (2) by taking the absolute value of all of the terms, incorporating Eq. (11), and using the fact that  $|\langle e^{i\nu t} \rangle_{L_T}| = e^{-|\nu|T}$ , we get

$$\langle \tilde{D}_A(\rho_t, \omega) \rangle_T \leq \frac{5\pi Q^2}{16} \sum_{\alpha\beta} p_\alpha p_\beta e^{-|G_\alpha - G_\beta|T}, \quad (12)$$

with  $p_\alpha \equiv (|v_\alpha|/Q)$  and  $Q \equiv \sum_\alpha |v_\alpha|$ .

We are interested in the decay in time of  $\langle \tilde{D}_A(\rho_t, \omega) \rangle_T$ . For the normalized probability distribution  $p_\alpha$ , we define the function  $\xi_p$  as follows.

**Definition 3** Given any normalized probability distribution  $p$  over the values of a real variable  $Y$ , we define  $\xi_p(x)$  as the maximum probability of any interval of length  $x$ . In particular, when  $Y$  is discrete,

$$\xi_p(x) = \max_{y_0 \in \mathbb{R}} \sum_{\alpha: y_\alpha \in [y_0, y_0+x]} p_\alpha. \quad (13)$$

In Appendix B, we prove the following.

**Proposition 4** (General bound). For any initial state  $\rho_0$ , any Hamiltonian, and any observable  $A$ ,

$$\langle \tilde{D}_A(\rho_t, \omega) \rangle_T \leq \pi Q^2 \xi_p \left( \frac{1}{T} \right). \quad (14)$$

The function  $\xi_p(x)$  will, in general, be difficult to compute explicitly, but for small  $x$ , it can be bounded

(and well approximated) by a linear function. We will capture this behavior in the following.

**Proposition 5** For any distribution  $p$ ,

$$\xi_p(x) \leq \frac{\xi_p(\epsilon)}{\epsilon} x + \xi_p(\epsilon), \quad \forall \epsilon \in (0, \infty). \quad (15)$$

It will be convenient to reexpress this as

$$\xi_p(x) \leq \frac{a(\epsilon)}{\sigma} x + \delta(\epsilon), \quad (16)$$

where  $\sigma$  is the standard deviation of the distribution, and we define

$$a(\epsilon) = \frac{\xi_p(\epsilon)}{\epsilon} \sigma, \quad \delta(\epsilon) = \xi_p(\epsilon). \quad (17)$$

*Proof.*—Take  $(n-1)\epsilon \leq x < n\epsilon$ , with  $n \geq 1$  a natural number. The function  $\xi_p$  is nondecreasing; hence,  $\xi_p(x) \leq \xi_p(n\epsilon)$ . Since  $\xi_p(\epsilon)$  quantifies the maximum probability that can fit *any* interval  $\epsilon$ , we also have that  $\xi_p(n\epsilon) \leq n\xi_p(\epsilon)$ , which results in

$$\xi_p(x) \leq \xi_p(n\epsilon) \leq (n-1)\xi_p(\epsilon) + \xi_p(\epsilon) \leq \frac{\xi_p(\epsilon)}{\epsilon} x + \xi_p(\epsilon). \quad (18)$$

□

We now derive some general properties of  $\xi_p$ . For many distributions  $p$ , we would expect to be able to find an  $\epsilon$  such that  $a(\epsilon) \sim 1$  (in terms of its approximate order of magnitude) and  $\delta(\epsilon) \ll 1$ . To visualize how this can be so, consider the case in which the distribution essentially has a single “peak” and in which the standard deviation  $\sigma$  approximately quantifies the width of this peak. In such a case, a rough estimate for the maximum probability that can fit inside an interval  $\epsilon$  can be given by

$$\xi_p(\epsilon) \sim \frac{\epsilon}{\sigma}. \quad (19)$$

With this estimate, we indeed get  $a(\epsilon) \sim 1$ . Figure 1 illustrates this for the case of a binomial distribution, where  $0.2 < a(\epsilon) < 0.8$  for all  $\epsilon > \frac{1}{2}$ .

In general, the above will work when the distribution  $p$  is approximately unimodal, i.e., characterized by a single distinct peak. If, on the contrary, the distribution was composed of two or more peaks, the estimate in Eq. (19) might not hold, as Fig. 2 exemplifies.

When Eq. (19) holds, taking  $\epsilon \ll \sigma$  is also enough to ensure  $\delta(\epsilon) \ll 1$ . Note that in Fig. 1,  $a(\epsilon)$  diverges for small  $\epsilon$ . To avoid such behavior, we would typically want to choose  $\epsilon$  larger than the gaps between consecutive values of the variable. Overall, we would expect to be able to find an



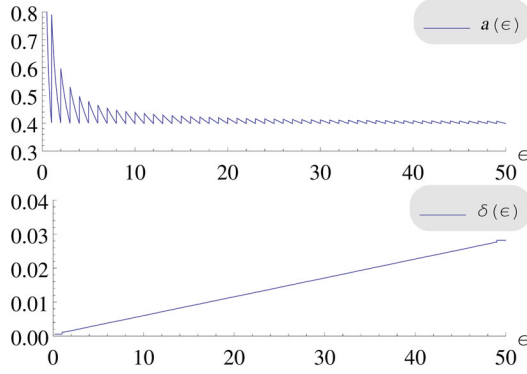


FIG. 1. We show  $a$  and  $\delta$  as functions of  $\epsilon$  for a binomial distribution of  $2 \times 10^6$  randomly chosen bits (mean  $10^6$ , standard deviation  $\approx 707$ ).

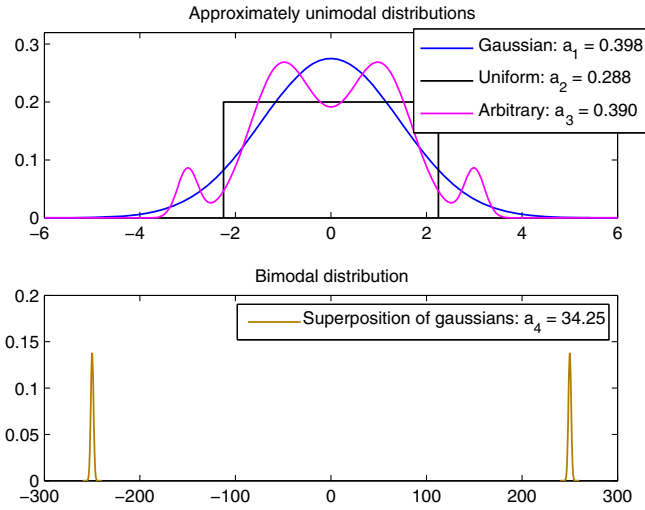


FIG. 2. Examples of approximately unimodal distributions (above) and bimodal distribution (below), for continuous distributions in the limit  $\epsilon \rightarrow 0$ . The bimodal distribution can violate the estimate  $\xi_p(\epsilon) \sim (\epsilon/\sigma)$ , simply because one can make the standard deviation arbitrarily large by placing the peaks further apart without changing the actual value of  $\xi_p(\epsilon)$ .

$\epsilon$  satisfying both  $a(\epsilon) \sim 1$  and  $\delta(\epsilon) \ll 1$  if the distribution is approximately unimodal and spread over many different values of the random variable.

In our particular case, Proposition 4 refers to the distribution  $p_\alpha = p_{(j,k)} = [(|\rho_{jk}| |A_{kj}|) / Q \|A\|]$ , which depends strongly on the distribution of energy gaps of the system. For large systems with typical energy ranges (e.g., finite positive temperatures), their energy levels tend to be more densely packed for larger energies, which leads to a much larger concentration of small gaps than large gaps. For most  $A$  and  $\rho_0$ , we would therefore expect the distribution  $p_\alpha$  to be more peaked towards the center and decay for larger values of the energy gaps  $G$ , leading to an approximately unimodal distribution over a dense spectrum as considered above. Nevertheless, this will not always be the case, as we will discuss in Sec. VI.

#### IV. OBSERVABLE-DEPENDENT TIME-SCALE BOUND

Propositions 4 and 5 lead to the following result.

**Theorem 6** (Observable-dependent bound). Given an initial state  $\rho_0$ , observable  $A$ , Hamiltonian  $H$ , and any  $\epsilon > 0$ , the time-averaged weak distinguishability satisfies

$$\begin{aligned} \langle \tilde{D}_A(\rho_t, \omega) \rangle_T &\leq \pi Q^2 \left( \frac{a(\epsilon)}{\sigma_G T} + \delta(\epsilon) \right) \\ &\leq \frac{\pi a(\epsilon) \|A\|^{1/2} Q^{5/2}}{T \sqrt{|\text{Tr}([\rho_0, H], H[A])|}} + c\delta(\epsilon) Q^2, \end{aligned} \quad (20)$$

where

$$Q^2 \leq d \text{Tr}[\rho_0^2], \quad (21)$$

$\sigma_G$  is the standard deviation of energy gaps  $G_\alpha$  for the distribution  $p_\alpha$ ,  $a(\epsilon)$  and  $\delta(\epsilon)$  are as in Proposition 5, and  $d$  is the rank of  $\omega$ .

Since the distribution  $p_\alpha$  is symmetric with respect to interchanging the indices  $\{j, k\}$  while  $G_\alpha$  is antisymmetric, we get that its variance, denoted by  $\sigma_G^2$ , satisfies

$$\begin{aligned} \sigma_G^2 &= \sum_\alpha p_\alpha G_\alpha^2 - \left( \sum_\alpha p_\alpha G_\alpha \right)^2 = \sum_\alpha p_\alpha G_\alpha^2 \\ &= \sum_{jk} \frac{|\rho_{jk}| |A_{kj}|}{Q \|A\|} (E_j - E_k)^2 \geq \frac{1}{Q \|A\|} \left| \sum_{jk} \rho_{jk} A_{kj} (E_j - E_k)^2 \right| \\ &= \frac{1}{Q \|A\|} |\text{Tr}([\rho_0, H], H[A])|. \end{aligned} \quad (22)$$

Notice that for a local Hamiltonian and observable, and a known initial state, this expression (combined with the bound for  $Q$  which soon follows) is much simpler to compute than  $\sigma_G$  since it does not require detailed knowledge of the Hamiltonian's spectrum and eigenbasis, which is needed in order to construct the distribution  $p_\alpha$  in the first place.

Moreover, we find an upper bound for  $Q$ ,

$$\begin{aligned} Q^2 &= \left( \sum_{jk: E_j \neq E_k} |\rho_{jk}| \frac{|A_{kj}|}{\|A\|} \right)^2 \\ &\leq \left( \sum_{jk} |\rho_{jk}| \frac{|A_{kj}|}{\|A\|} \right)^2 \leq \left( \sum_{jk} |\rho_{jk}|^2 \right) \left( \sum_{j'k' \in \text{supp}(\omega)} \frac{|A_{j'k'}|^2}{\|A\|^2} \right) \\ &= \text{Tr}[\rho_0^2] \frac{\text{Tr}[\Pi_\omega A \Pi_\omega A]}{\|A\|^2} \leq \text{Tr}[\rho_0^2] \text{Tr}[\Pi_\omega] = d \text{Tr}[\rho_0^2], \end{aligned} \quad (23)$$

where  $\Pi_\omega$  projects onto all the energy levels of the Hamiltonian that occur with nonzero probability in  $\rho_0$  (this is given by the support of  $\omega$ ). In the second line, we restrict ourselves to this set of energy levels and use the Cauchy-Schwarz inequality. For the last line, notice that using

the Cauchy-Schwarz inequality for the scalar product  $(O, P) \equiv \text{Tr}[OP^\dagger]$ , we can bound  $\text{Tr}[\Pi_\omega A \Pi_\omega A] \leq \text{Tr}[\Pi_\omega A^2]$ . Then, using that for any two positive semidefinite matrices,  $\text{Tr}[OP] \leq \|O\| \text{Tr}[P]$ , we find  $\text{Tr}[\Pi_\omega A^2] \leq \text{Tr}[\Pi_\omega] \|A\|^2 = d \|A\|^2$ .

Inserting the above into Eqs. (14) and (16) proves our claim.  $\square$

If the second term on the right-hand side of Eq. (20) is small, the system will eventually equilibrate with respect to  $A$ . The time dependence is determined by the first term. In particular, the system will be equilibrated (in the sense described in Sec. II) for times  $T \gg T_{\text{eq}}$ , where

$$T_{\text{eq}} \equiv \frac{\pi a(\epsilon) \|A\|^{1/2} Q^{5/2}}{\sqrt{|\text{Tr}([\rho_0, H], H|A)]|}}. \quad (24)$$

It is interesting to note the dependence of the above expression on  $[\rho_0, H]$ , which is, up to a minus sign, the second time derivative of the state at  $t = 0$ . Therefore,  $T_{\text{eq}}$  can be alternatively written as

$$T_{\text{eq}} = \frac{\pi a(\epsilon) \|A\|^{1/2} Q^{5/2}}{\sqrt{|\text{Tr}[\frac{d^2 \rho_t}{dt^2}|_{t=0} A]|}}. \quad (25)$$

Remarkably, the denominator of this expression is what one would expect from a Taylor expansion of the distance for short times, assuming the system is initially as far from equilibrium as possible (then, the first derivative term is 0, and one is left with the second derivative as leading order).

We argued earlier that we would typically expect  $a(\epsilon) \sim 1$  and  $\delta(\epsilon) \ll 1$ . However, we still have to address the size of the bound for  $Q$  given by Eq. (21), which could greatly influence the speed of equilibration. Notice that, in general, the dimension  $d$  of the Hilbert space is extremely large since it scales exponentially with the number of constituents of the closed system being considered. Therefore, in order for this bound to show rapid equilibration, we would need a very mixed initial state, spread over a significant fraction of the Hilbert space.

Moreover, the constant  $Q$  appears in the second term in Theorem 6, along with  $\delta(\epsilon)$ . In order to show equilibration at all, this second term needs to be small, too.

In the next section, we consider an important physical scenario and then use our bound to show reasonably fast equilibration.

## V. SYSTEM INTERACTING WITH A BATH

We now turn to the paradigmatic case of a small system interacting with a large thermal bath. This situation corresponds to decomposing the closed system considered in the previous sections into a small system  $\mathcal{S}$  and a bath  $\mathcal{B}$ . By assuming the observable  $A$  to be of the form

$$A = A_S \otimes \mathbb{1}_B, \quad (26)$$

where  $A_S$  acts on the system and  $\mathbb{1}_B$  is the identity acting on the bath, one can focus on the system's behavior.

The total Hamiltonian is denoted by

$$H = H_S + H_B + H_I, \quad (27)$$

where  $H_S$  and  $H_B$  are the system and bath Hamiltonians, and  $H_I$  denotes the interaction between them.

We assume that the system  $\mathcal{S}$  is initially in an arbitrary state  $\rho_S$  and, for simplicity, not correlated with the initial state of bath  $\rho_B$ , that is,  $\rho_0 = \rho_S \otimes \rho_B$ , corresponding to a system that is initially isolated and suddenly allowed to interact with  $\mathcal{B}$  via  $H_I$ .

To show that such a situation can lead to a small value for  $Q$ , we first consider the case in which the bath is in a maximally mixed state, with

$$\rho_0 = \rho_S \otimes \frac{\mathbb{1}_B}{d_B}. \quad (28)$$

In this case, it is easy to see from Eq. (21) that

$$Q \leq d \text{Tr}[\rho_0^2] \leq (d_S d_B) \text{Tr}_S[\rho_S^2] \text{Tr}_B\left[\frac{\mathbb{1}_B^2}{d_B^2}\right] = d_S \text{Tr}_S[\rho_S^2]. \quad (29)$$

The remainder of this section corresponds to extending this simple example (which could be understood as a system interacting with an infinite-temperature bath) to the more physical case of a system interacting with a finite-temperature bath.

In what follows, given a Hamiltonian  $H$ , we denote an energy window of width  $\Delta$  centered around an energy  $E$  in terms of its corresponding Hilbert space  $\mathcal{H}_H^{E,\Delta}$ , defined as

$$\mathcal{H}_H^{E,\Delta} = \text{span}\left\{|E_n\rangle : E - \frac{\Delta}{2} \leq E_n \leq E + \frac{\Delta}{2}\right\}. \quad (30)$$

We consider the state of the bath from the microcanonical ensemble viewpoint. Consequently, we consider an energy window of the bath Hamiltonian of width  $\Delta$  centered around  $E_B$ . The subspace that this defines,  $\mathcal{H}_{H_B}^{E_B,\Delta}$ , will be referred to as a microcanonical window, and its dimension will be denoted by  $d_B^\Delta$ . The initial state of the bath is then  $\rho_B = [(\mathbb{1}_B^\Delta)/d_B^\Delta]$ , and the initial state of the system plus bath is

$$\rho_0 = \rho_S \otimes \frac{\mathbb{1}_B^\Delta}{d_B^\Delta}, \quad (31)$$

which we will use from now on.

The width of the microcanonical window is to be taken large enough such that it contains many energy levels, in particular, many more than the dimension of the system, yet small in comparison to the whole spectrum of the bath Hamiltonian  $H_B$ .

### A. Truncation of the Hilbert space

Notice that the state  $\rho_0 = \rho_S \otimes [(\mathbb{1}_B^\Delta)/d_B^\Delta]$ , corresponding to a bath in the microcanonical ensemble, is quite mixed. This is good news for our bound  $Q^2 \leq d\text{Tr}[\rho_0^2]$  given by Eq. (21) since the purity of the state will be a small number. However, the presence of the dimension of the Hilbert space implies that the bound for  $Q$  could still be extremely large. In this section, we show a truncation method for the state and the Hilbert space, which allows us to reduce the relevant dimension significantly.

As the eigenvalues of  $H_S$  lie between  $-||H_S||$  and  $||H_S||$ , the initial state  $\rho_0$  is contained inside an energy window of width  $\Delta + 2||H_S||$  of the unperturbed Hamiltonian  $H_S + H_B$ . In other words,  $\rho_0$  lies within the subspace  $\mathcal{H}_{H_B+H_S}^{E_B, \Delta+2||H_S||}$ .

However, this no longer holds when considering the full Hamiltonian  $H = H_S + H_B + H_I$ ; in principle,  $\rho_0$  can have support outside the subspace  $\mathcal{H}_{H_B+H_S+H_I}^{E_B, \Delta+2||H_S||}$ . Yet, one has the intuition that if the interaction is small compared to the unperturbed Hamiltonian, the energy window where the state is supported should not grow significantly. This would imply that one does not need to consider the full Hilbert space but rather a truncated subspace corresponding to a window somewhat larger than the original  $\mathcal{H}_{H_B+H_S}^{E_B, \Delta+2||H_S||}$ .

The above reasoning is proved to be correct in Appendix C, where we show that the trace distance between the state  $\rho_0$  and a truncated state  $\Pi\rho_0\Pi$  is small, where  $\Pi$  is a projector onto the truncated subspace. More precisely, we find the following.

**Proposition 7** (Hilbert space truncation). For any  $K$ , the state  $\rho_0 = \rho_S \otimes [(\mathbb{1}_B^\Delta)/d_B^\Delta]$  that lies within  $\mathcal{H}_{H_B+H_S}^{E_B, \Delta+2||H_S||}$  can be truncated to the state  $\Pi\rho_0\Pi$ , with

$$||\rho_0 - \Pi\rho_0\Pi||_1 \leq \frac{2}{K}, \quad (32)$$

where  $\Pi$  projects onto the subspace  $\mathcal{H}_{H_B+H_S+H_I}^{E_B, \Delta+2||H_S||+\eta}$  with a width extended by  $\eta = \sqrt{8d_S}||H_I||K$ .

As a straightforward corollary, one obtains that

$$\tilde{\mathcal{D}}_A(\rho_t, \Pi\rho_t\Pi) \leq \frac{1}{K^2}, \quad (33)$$

where  $\rho_t = e^{-iHt}\rho_0 e^{iHt}$  is the evolved state. This shows that, as long as we take  $K$  large enough, the two states give similar evolutions.

This truncation procedure will be particularly useful to us since the dimension of the accessible Hilbert space  $\mathcal{H}_{H_B+H_S+H_I}^{E_B, \Delta+2||H_S||+\eta}$  is, in general, much smaller than the full dimension.

We also find in Appendix C that, if the density of states of the bath is denoted by  $\nu_B(E)$ , the dimension of the truncated state,  $d_{\text{trunc}} = \text{rank}(\Pi)$ , satisfies

$$d_{\text{trunc}} \leq \frac{d_S}{1 - \frac{1}{K}} \int_{E_B - (\Delta/2) - (1 + \sqrt{2d_S})K||H_I|| - ||H_S||}^{E_B + (\Delta/2) + (1 + \sqrt{2d_S})K||H_I|| + ||H_S||} \nu_B(E) dE. \quad (34)$$

Meanwhile, the dimension of the (unperturbed) microcanonical window of the bath is given by

$$d_B^\Delta = \int_{E_B - (\Delta/2)}^{E_B + (\Delta/2)} \nu_B(E) dE. \quad (35)$$

Typically, thermal baths have a (coarse-grained) density of states which grows approximately exponentially with energy. Thus, if we take

$$\nu_B(E) = \mathcal{N} e^{\beta E}, \quad (36)$$

where  $\beta$  is the inverse temperature and  $\mathcal{N}$  a normalization constant, it is easy to obtain

$$\begin{aligned} d_{\text{trunc}} &\leq \frac{d_S d_B^\Delta}{(1 - \frac{1}{K})} \frac{\sinh[\beta \frac{\Delta}{2} + \beta ||H_S|| + (1 + \sqrt{2d_S})K\beta ||H_I||]}{\sinh[\beta \frac{\Delta}{2}]} \\ &\leq \frac{d_S d_B^\Delta}{(1 - \frac{1}{K})(1 - e^{-\beta\Delta})} e^{\beta ||H_S|| + (1 + \sqrt{2d_S})K\beta ||H_I||}. \end{aligned} \quad (37)$$

Note that, given that the energy width of the microcanonical window grows as the number of constituents of the bath increases, in general,  $\beta\Delta \gg 1$  holds for a large enough bath, in which case the last inequality is a particularly good approximation.

### B. Time scales for a system in contact with a bath

Proposition 7 allows us to truncate the microcanonical state  $\rho_0$  to  $\Pi\rho_0\Pi$  since the error introduced is small. This greatly reduces the dimension of the relevant Hilbert space and, consequently, the corresponding bound for the constant  $Q$  in Theorem 6.

However, this reasoning would also lead us to use the truncated state in the theorem itself, which would cause the replacement of  $\text{Tr}([\rho_0, H], H)A$  by  $\text{Tr}([\Pi\rho_0\Pi, H], H)A$  in Eq. (20). This not only introduces additional complexity but could possibly significantly weaken the bound. Moreover, even if the Hamiltonian involved nearest-neighbor-type interactions,  $\Pi$  could be highly nonlocal, and indeed, we may have no way of computing it. Nevertheless, we prove in Appendix D that the time average of the weak distinguishability can be bounded with a commutator involving the original state  $\rho_0$  instead of the truncated one while still having a relevant Hilbert space with much smaller dimension than the original space.

We finally have all the ingredients to apply Theorem 6 to the case of a system in contact with a thermal bath, which turns into the following.



**Theorem 8** (Bound for a system interacting with a thermal bath). For any  $\epsilon > 0$ ,  $K > 0$ , observable  $A$ , Hamiltonian  $H = H_B + H_S + H_I$ , and initial state  $\rho_0 = \rho_S \otimes [(\mathbb{1}_B^\Delta)/d_B^\Delta]$ , the weak distinguishability satisfies

$$\langle \tilde{D}_A(\rho_t, \omega) \rangle_T \leq \frac{\pi a(\epsilon) \|A\|^{1/2} Q_2^{5/2}}{T \sqrt{|\text{Tr}([\rho_0, H], H[A])|}} + \pi \delta(\epsilon) Q_2^2 + \frac{18}{K^2}, \quad (38)$$

where  $a(\epsilon)$  and  $\delta(\epsilon)$  are as in Proposition 5, and

$$Q_2 \leq \sqrt{\text{Tr}[\rho_0^2] d_{\text{trunc}}} + \frac{2}{K},$$

$$d_{\text{trunc}} \leq \frac{d_S}{1 - \frac{1}{K}} \int_{E_B - \frac{\Delta}{2} - (1 + \sqrt{2d_S}K\|H_I\| - \|H_S\|)}^{E_B + \frac{\Delta}{2} + (1 + \sqrt{2d_S}K\|H_I\| + \|H_S\|)} \nu_B(E) dE, \quad (39)$$

where  $\nu_B(E)$  is the density of states of the bath. Moreover, if we take the density of states of the bath to be  $\nu_B(E) \propto e^{\beta E}$  (as we would expect for a thermal bath in the vicinity of the microcanonical window), we obtain from Eq. (37) that

$$Q_2 \leq \sqrt{\frac{d_S \text{Tr}_S[\rho_S^2] e^{\beta \|H_S\| + (1 + \sqrt{2d_S}K\beta\|H_I\|)}}{(1 - \frac{1}{K})(1 - e^{-\beta\Delta})}} + \frac{2}{K}. \quad (40)$$

Taking a system observable  $A = A_S \otimes \mathbb{1}_B$ , we recover the main result in Sec. II since the Hamiltonian  $H_B$  in Eq. (38) commutes with  $\rho_0$  and  $A$ .

Let us consider this result more closely. To begin with, all time-independent terms have to be small for our theorem to imply equilibration in the first place. The factors involving  $K$  in Eqs. (38) and (40) come from the truncation procedure and are small as long as the microcanonical window and the truncation window are large enough.

The other time-independent term is  $\pi \delta(\epsilon) Q_2^2$ , which we have neglected so far. As discussed in Sec. III, for distributions  $p_\alpha$  that are approximately unimodal and sufficiently spread over different values, one can estimate that  $\delta(\epsilon) \sim [\epsilon/(\sigma_G)]$ . Notice that as the bath grows in size, one would expect that this holds for smaller values of  $\epsilon$  since the distribution  $p_\alpha$  would be spread over more values. We could therefore take  $\epsilon$  smaller and smaller and reduce  $\delta(\epsilon)$ . At the same time, the bound on  $Q_2$  in Eq. (40) will generally not grow with the dimension of the bath. To see this, note that typically (e.g., for short-ranged interactions in a lattice system),  $\|H_I\|$  will not increase significantly as the bath size increases and that increasing the width  $\Delta$  of the microcanonical window as the bath grows will cause the bound to become tighter. Therefore, in the limit of increasing bath sizes, the term  $\pi \delta(\epsilon) Q_2^2$  becomes negligible, as needed.

The fact that the results in Theorem 8 do not depend on the dimension of the full Hilbert space is a very noticeable aspect of this paper. This is in stark contrast with previously known general upper bounds on the time scale of equilibration [1], which essentially scale with the full Hilbert space dimension.

Finally, the first term in Eq. (38) determines the time decay of the weak distinguishability and can be interpreted the same way as in the corresponding term in Theorem 6 (see subsequent discussion). Notice that, once  $a(\epsilon)$  is estimated, the time dependence can, in general, be calculated analytically for a given initial state, Hamiltonian, and observable. Moreover, performing this calculation is much simpler than solving the exact time evolution, which involves commutators of the initial state and Hamiltonian of all orders and can only be done for simple models.

It is illuminating to ask how our bound behaves in a case where no equilibration occurs. Take, for example, a spin 1/2 in a pure initial state  $|\Psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$  as the system  $S$ , and a bath composed of  $N$  other spins in the microcanonical ensemble. Furthermore, take the Hamiltonian  $H = \Omega \sigma_z^S + H_B$  and the observable  $A = \sigma_x^S \otimes \mathbb{1}_B$ . Since the system does not interact with the bath, it does not equilibrate with respect to the observable  $A$ . The key to understanding where our bound expresses this fact is in the factor  $\delta(\epsilon)$ . It is easy to see that the distribution  $p_\alpha$  is composed of only two values, corresponding to the gaps  $\Omega$  and  $-\Omega$ , which results in  $\delta(\epsilon) \geq \frac{1}{2}$  for any  $\epsilon$  and hence no equilibration at all.

### C. System interacting with an environment in a pure state: Typical behavior

So far, we have considered mixed initial states of the total closed system. Here, we show that our results can be extended to the typical behavior of pure initial states of the environment that interact with the system.

Let us consider the environment's initial state to be pure and drawn at random from the microcanonical window. Any pure state from the microcanonical window can be written as

$$\rho_B^U = U|\psi\rangle\langle\psi|U^\dagger, \quad (41)$$

where  $U$  is a unitary operator acting on  $\mathcal{H}_{H_B}^{E_B, \Delta}$ . By averaging over all possible  $U$ 's, drawn from the Haar measure, we have the typical behavior for random pure states from the microcanonical subspace  $\mathcal{H}_{H_B}^{E_B, \Delta}$ .

It turns out that taking the initial environment state to be a pure state chosen at random from a microcanonical window leads to results very similar to the environment starting in the microcanonical mixed state  $[(\mathbb{1}_B^\Delta)/d_B^\Delta]$  in the subspace  $\mathcal{H}_{H_B}^{E_B, \Delta}$ . More precisely, we show, in Appendix E, the following.

**Proposition 9** (Evolution for typical initial states of the bath). The weak distinguishability averaged over all

possible initial pure states of the environment drawn from a microcanonical window of width  $\Delta$  satisfies

$$\langle \tilde{\mathcal{D}}_A(\rho_t^U, \omega^U) \rangle_U \leq \tilde{\mathcal{D}}_A(\rho_t, \omega) + \frac{d_S}{d_B^\Delta}, \quad (42)$$

where  $\rho_0^U = \rho_S \otimes U|\psi\rangle\langle\psi|U^\dagger$  with corresponding evolved and equilibrium states  $\rho_t^U$  and  $\omega^U$ , and  $\rho_0 = \rho_S \otimes [(\mathbb{1}_B^\Delta)/d_B^\Delta]$  with corresponding evolved and equilibrium states  $\rho_t$  and  $\omega$ .

Since the microcanonical window is assumed to contain many more levels than the system's dimension,  $d_B^\Delta \gg d_S$ , the above expression implies that, for typical initial pure states of the bath, the evolution is as if the initial state was  $\rho_0 = \rho_S \otimes [(\mathbb{1}_B^\Delta)/d_B^\Delta]$ .

It is straightforward to combine this Proposition with Theorem 8 and show that the upper bound for the typical time scale of equilibration for a system interacting with a bath in a pure state is the same as if the bath were in the microcanonical state.

## VI. DISCUSSION

From previous work, we know that one needs to impose further conditions in order to prove reasonably fast equilibration since extremely slow observables can always be constructed [2,3].

In this article, we have found a set of sufficient conditions that ensure this. More precisely, when the distribution  $p_\alpha \equiv [1/(Q\|A\|)]|\rho_{jk}||A_{kj}|$  (which characterizes the energy gaps that are most relevant to the particular state and observable under consideration) is approximately unimodal and spread over many different values, one expects  $a(\epsilon) \sim 1$  and  $\delta(\epsilon) \ll 1$ . In the setting of a system interacting with a thermal bath, this implies equilibration time scales that do not scale as the size of the bath grows, as Theorem 8 shows.

Whether the above holds or not ultimately boils down to the values of the off-diagonal matrix elements of the observable and initial state in the energy basis, and to the distribution of energy gaps. Nevertheless, there are general arguments indicating that  $a(\epsilon) \sim 1$  and  $\delta(\epsilon) \ll 1$  might hold for a wide range of systems. First, in typical situations, one might expect that a state is spread roughly equally over a range of energies (this occurs, for example, for thermal states) and that, unless the observable  $A$  is fine-tuned, its components are also spread relatively smoothly over this band of energies, at least in a coarse-grained sense. Second, for large systems, the distribution of energy levels tends to grow exponentially with energy in the region of finite temperature. It is easy to check that if one assumes a density of states  $\nu(E) \propto e^{\beta E}$ , the corresponding density of gaps scales like  $\mu(|G|) \propto e^{-\beta|G|}$ , with an exponential decrease. This implies that, in order to have a resulting distribution that is characterized by one peak, it is *sufficient* to have matrix elements of  $A$  and of the initial state  $\rho_0$  that

grow subexponentially as a function of the energy gaps. This does not seem like a particularly strong assumption. Finally, let us note that even when  $p_\alpha$  is not unimodal and is instead composed of a number of distinct peaks, Theorem 8 will still give reasonable equilibration times (in particular, times that are approximately independent of the size of the bath) except in the case where the individual peaks get sharper as the size of the bath is increased. Our intuition is that such behavior is rare and thus that the bound will have very general applicability.

The remaining question is whether physically relevant cases will, in general, be of this form [satisfying  $a(\epsilon) \sim 1$  and  $\delta(\epsilon) \ll 1$ , and therefore “reasonably fast equilibrating”] or of the other (violating these conditions and therefore “slow equilibrating”). Appendix F illustrates the transition to approximate unimodality, and the conditions being met, as environment size increases in a simulation of a 1D Ising model with transversal magnetic field and periodic boundary conditions. However, proving that this occurs and finding the physical conditions under which it happens remain interesting open problems for future study.

It is worthwhile comparing our conditions with the assumptions made in previous work in order to prove equilibration of closed quantum systems. Equilibration can be proven by assuming that the effective dimension defined as  $d_{\text{eff}} = [1/(\sum_j \rho_{jj}^2)]$  (for a nondegenerate spectrum, for simplicity) is large [25,31] and that the Hamiltonian does not have too many degenerate energy gaps [1]. Notice that, although we do not make these assumptions explicitly, we are, in some sense, implicitly assuming both of them. On the one hand, a high effective dimension is related to having many energy levels populated in the system, which is necessary in order to have a distribution  $p_\alpha$  that is spread over many different values. On the other hand, the presence of a very degenerate energy gap results in a distribution  $p_\alpha$  such that  $\delta(\epsilon) \ll 1$  does not hold, as the simple example after Theorem 8 illustrates.

The present paper emphasizes the importance of the off-diagonal matrix elements of the observable and initial state to the study of the equilibration time scales in closed quantum systems.

The eigenstate thermalization hypothesis, introduced by Deutsch and Srednicki as a sufficient condition for thermalization [33,34], has motivated extensive work on the distribution of diagonal matrix elements of observables [35–42]. However, much less work dealing with the distribution of the off-diagonal matrix elements is available, some examples being Refs. [43–46]. The recent papers [44,45] show, in certain models, Gaussian distributions of these matrix elements for local observables, which supports our claim that a unimodal distribution of  $p_\alpha$  is to be expected in many situations. Moreover, Ref. [46] numerically verifies our predictions in an experimentally realizable setup consisting of an electron in a quantum dot interacting with a bath of nuclear spins. Remarkably, the

authors find that, even though our results are model independent and not tailored to this particular system, our new bounds fall within 2 orders of magnitude of the actual time scale.

Our paper focuses on the equilibration of a small system with respect to a pre-equilibrated bath, but many open questions remain regarding general equilibration time scales. One direction of particular interest is an equilibration time scale for the bath itself, as well as what aspects are necessary for it to play its usual thermodynamic role. We hope that the tools developed here will aid in further study along these lines and help shed more light into this important topic.

## ACKNOWLEDGMENTS

We would like to thank Moritz Fuchs, Daniel Hetterich, and Björn Trauzettel for many illuminating discussions, and for sharing the findings of Ref. [46] prior to publication. Part of this work was supported by the COST Action MP1209 “Thermodynamics in the Quantum Regime.” A. S. L. M. acknowledges support from the CNPq. A. J. S. acknowledges support from the Royal Society. A. W. is supported by the EU (STREP “RAQUEL”), the ERC (AdG “IRQUAT”), the Spanish MINECO (Grant No. FIS2013-40627-P) with the support of FEDER funds, as well as by the Generalitat de Catalunya CIRIT, Project No. 2014-SGR-966.

## APPENDIX A: EQUILIBRATION IN TERMS OF DISTINGUISHABILITY

Equilibration of the expectation value of some observable does not imply equilibration of the observable itself. Here, we show how the results can be cast into a stronger sense of equilibration in terms of the distinguishability, as we used in Ref. [3]. Distinguishability of states  $\rho$  and  $\sigma$  with respect to an observable  $\mathcal{M} = \{P_1, P_2, \dots, P_N\}$ , where  $P_i$  are a complete set of projectors, is defined by

$$\mathcal{D}_{\mathcal{M}}(\rho, \sigma) = \frac{1}{2} \sum_i |\text{Tr}[P_i \rho] - \text{Tr}[P_i \sigma]|, \quad (\text{A1})$$

and it characterizes the probability of successfully guessing between the two states (assuming they are given with equal probabilities), via  $p_{\text{success}} = \frac{1}{2} + \frac{1}{2} \mathcal{D}_{\mathcal{M}}(\rho, \sigma)$ . By Jensen’s and Cauchy-Schwarz’s inequalities, we can relate the distinguishability  $\mathcal{D}_{\mathcal{M}}$  to the weak distinguishability  $\tilde{\mathcal{D}}_{P_i}$  considered in this paper:

$$\begin{aligned} \langle \mathcal{D}_{\mathcal{M}}(\rho_t, \omega) \rangle_T &\leq \sqrt{\langle \mathcal{D}_{\mathcal{M}}^2(\rho_t, \omega) \rangle_T} \\ &\leq \sqrt{N} \sqrt{\frac{1}{4} \sum_i \langle |\text{Tr}[P_i \rho_t] - \text{Tr}[P_i \omega]|^2 \rangle_T} \\ &= \sqrt{N} \sqrt{\sum_i \langle \tilde{\mathcal{D}}_{P_i}(\rho_t, \omega) \rangle_T}. \end{aligned} \quad (\text{A2})$$

Each term  $\langle \tilde{\mathcal{D}}_{P_i}(\rho_t, \omega) \rangle_T$  can be bounded via the results from Theorems 6 and 8, and therefore, fast equilibration of the projectors will imply fast equilibration of the distinguishability.

## APPENDIX B: THE $\xi_p(\frac{1}{T})$ FUNCTION—PROPOSITION 4

We wish to bound the decay with time  $T$  of

$$\sum_{\alpha\beta} p_{\alpha} p_{\beta} e^{-|G_{\alpha}-G_{\beta}|T}. \quad (\text{B1})$$

To connect this to the function

$$\xi_p\left(\frac{1}{T}\right) = \max_{G \in \mathbb{R}} \sum_{\substack{\alpha: \\ G_{\alpha} \in [G, G+\frac{1}{T}]}} p_{\alpha}, \quad (\text{B2})$$

we define the auxiliary function

$$g(x) = \begin{cases} 1 & \text{if } x \in [0, 1) \\ 0 & \text{otherwise} \end{cases}. \quad (\text{B3})$$

This definition allows us to upper bound the exponential as

$$e^{-|x|} \leq \sum_{n=0}^{\infty} e^{-n} g(|x| - n). \quad (\text{B4})$$

One can then see

$$\begin{aligned} \sum_{\alpha\beta} p_{\alpha} p_{\beta} e^{-|G_{\alpha}-G_{\beta}|T} &\leq \sum_{n=0}^{\infty} e^{-n} \sum_{\alpha} p_{\alpha} \sum_{\beta} p_{\beta} g(|G_{\alpha}-G_{\beta}|T - n) \\ &= \sum_{n=0}^{\infty} e^{-n} \sum_{\alpha} p_{\alpha} \sum_{\substack{\beta: \\ (|G_{\alpha}-G_{\beta}|T-n) \in [0,1)}} p_{\beta} \\ &\leq \sum_{n=0}^{\infty} e^{-n} \sum_{\alpha} p_{\alpha} \left[ \sum_{\substack{\beta: \\ G_{\beta} \in I_{+}^{(n)}}} p_{\beta} + \sum_{\substack{\beta: \\ G_{\beta} \in I_{-}^{(n)}}} p_{\beta} \right] \\ &\leq \sum_{n=0}^{\infty} e^{-n} \sum_{\alpha} p_{\alpha} \left[ 2\xi_p\left(\frac{1}{T}\right) \right] \\ &= \frac{2}{1-e^{-1}} \xi_p\left(\frac{1}{T}\right), \end{aligned} \quad (\text{B5})$$

where  $I_{+}^{(n)} = [G_{\alpha} + (n/T), G_{\alpha} + [(n+1)/T])$  and  $I_{-}^{(n)} = (G_{\alpha} - [(n+1)/T], G_{\alpha} - (n/T)]$ , and the inequality in the penultimate line is valid for any  $n$  and  $\alpha$ .

Together with Eq. (12),

$$\langle \tilde{\mathcal{D}}_A(\rho_t, \omega) \rangle_T \leq \frac{5\pi Q^2}{16} \sum_{\alpha\beta} p_{\alpha} p_{\beta} e^{-|G_{\alpha}-G_{\beta}|T}, \quad (\text{B6})$$

we obtain

$$\begin{aligned} \langle \tilde{\mathcal{D}}_A(\rho_t, \omega) \rangle_T &\leq \frac{5\pi}{8(1-e^{-1})} Q^2 \xi_p \left( \frac{1}{T} \right) \\ &\leq \pi Q^2 \xi_p \left( \frac{1}{T} \right), \end{aligned} \quad (\text{B7})$$

showing Proposition 4.

### APPENDIX C: TRUNCATION OF THE HILBERT SPACE—PROPOSITION 7

As mentioned in the main text, the state  $\rho_0 = \rho_S \otimes (1_B^\Delta/d_B^\Delta)$  lies within the subspace  $\mathcal{H}_{H_S+H_B}^{E_B, \Delta+2\|H_S\|}$ . We will show that, when considering the full interacting Hamiltonian  $H = H_S + H_B + H_I$ , it is enough to consider the truncated subspace  $\mathcal{H}_{H_S+H_B+H_I}^{E_B, \Delta+2\|H_S\|+\eta}$ , where  $\eta$  denotes the amount by which the energy window is extended.

The effect of “cutting” the state outside the space  $\mathcal{H}_{H_S+H_B+H_I}^{E_B, \Delta+2\|H_S\|+\eta}$  will obviously introduce errors. We wish to do it such that the truncated state remains close to the original one in trace distance:

$$\|\rho_0 - \Pi \rho_0 \Pi\|_1 \leq \frac{2}{K}, \quad (\text{C1})$$

where  $\Pi$  is the projector onto  $\mathcal{H}_{H_S+H_B+H_I}^{E_B, \Delta+2\|H_S\|+\eta}$  and  $\eta = \sqrt{8d_S} \|H_I\| K$ .

An arbitrary initial state of the system can be written as

$$\rho_S = \sum_{l,l'} \rho_{ll'}^S |E_l^S\rangle \langle E_{l'}^S|, \quad (\text{C2})$$

where  $|E_l^S\rangle$  are eigenvectors of  $H_S$  with energy  $E_l^S$ . As seen in the main text, the state of the bath is effectively proportional to the identity on a window of  $H_B$  and therefore diagonal in the basis  $H_B$ . Hence, we can write the total state as

$$\begin{aligned} \rho_0 &= \sum_{\lambda} \sum_{l,l'} c_{\lambda} \rho_{ll'}^S |E_l^S\rangle \langle E_{l'}^S| \otimes |E_{\lambda}^B\rangle \langle E_{\lambda}^B| \\ &\equiv \sum_{\lambda, l, l'} c_{\lambda} \rho_{ll'}^S |E_l^S, E_{\lambda}^B\rangle \langle E_{l'}^S, E_{\lambda}^B|, \end{aligned} \quad (\text{C3})$$

where  $|E_{\lambda}^B\rangle$  are the eigenvectors of  $H_B$  and the coefficients  $c_{\lambda}$  are positive and normalized (we do the calculation for an arbitrary state of the bath commuting with  $H_B$ , but in our case, actually  $c_{\lambda} = 1/d_B^\Delta$ ).

The following result will be useful.

**Lemma 10** (Gentle measurement [47]). For any state  $\rho$  and positive operator  $X$  such that  $X \leq I$  and

$$1 - \text{Tr}[\rho X] \leq \frac{1}{2K^2} \leq 1, \quad (\text{C4})$$

one has

$$\|\rho - \sqrt{X} \rho \sqrt{X}\|_1 \leq \frac{2}{K}. \quad (\text{C5})$$

In order to apply the lemma for the state  $\rho = \rho_0$  and the operator  $X = \sqrt{X} = \Pi$ , we start with

$$\begin{aligned} 1 - \text{Tr}[\rho_0 \Pi] &= \text{Tr}[\rho_0 \Pi^\perp] \\ &= \sum_{\lambda, l, l'} c_{\lambda} \rho_{ll'}^S \text{Tr}[|E_l^S, E_{\lambda}^B\rangle \langle E_{l'}^S, E_{\lambda}^B| \Pi^\perp] \\ &\leq \sum_{\lambda, l, l'} c_{\lambda} |\rho_{ll'}^S| |\langle E_{l'}^S, E_{\lambda}^B| \Pi^\perp |E_l^S, E_{\lambda}^B\rangle| \\ &\leq \sum_{\lambda, l, l'} c_{\lambda} |\rho_{ll'}^S| \|\Pi^\perp |E_l^S, E_{\lambda}^B\rangle\| \|\Pi^\perp |E_{l'}^S, E_{\lambda}^B\rangle\|, \end{aligned} \quad (\text{C6})$$

by taking absolute values in line 3, using the Cauchy-Schwarz inequality (with  $\|\psi\| = \sqrt{\langle \psi | \psi \rangle}$  as usual) in line 4, and denoting the orthogonal complement of  $\Pi$  by  $\Pi^\perp$ .

In order to upper bound this expression, we use Bhatia’s perturbation theory result (Theorem VII.3.1) in Ref. [48] (for another very interesting application treating the problem of proving thermalization in closed quantum systems, see Ref. [49]).

**Theorem 11** Let  $O$  and  $P$  be normal operators, let  $S_1$  and  $S_2$  be two subsets of the complex plane that are separated by a strip (or annulus) of width  $\Delta$ , and let  $E$  ( $F$ ) denote the orthogonal projection onto the subspace spanned by the eigenvectors of  $O$  ( $P$ ) corresponding to those of its eigenvalues that lie in  $S_1$  ( $S_2$ ). Then, for every unitarily invariant norm  $||| \cdot |||$ ,

$$|||EF||| \leq \frac{|||O - P|||}{\Delta}. \quad (\text{C7})$$

In our notation, this theorem implies

$$\|\Pi^\perp |E_l^S, E_{\lambda}^B\rangle\| = \|\Pi^\perp |E_l^S, E_{\lambda}^B\rangle \langle E_{l'}^S, E_{\lambda}^B|\| \leq \frac{\|H_I\|}{\Delta_{l,\lambda}}, \quad (\text{C8})$$

where we have related the Euclidean vector norm on the left to the operator norm on the right. In the above expression,  $\Delta_{l,\lambda}$  is the distance between the supports of  $|E_l^S, E_{\lambda}^B\rangle \langle E_{l'}^S, E_{\lambda}^B|$  and  $\Pi^\perp$ . Note that this distance satisfies  $\Delta_{l,\lambda} \geq (\eta/2)$ .

Using Eqs. (C6) and (C8), the fact that

$$\left( \sum_{l,l'} |\rho_{ll'}^S| \right)^2 \leq d_S^2 \sum_{l,l'} |\rho_{ll'}^S|^2 = d_S^2 \text{Tr}[\rho_S^2] \leq d_S^2, \quad (\text{C9})$$

and  $\sum_{\lambda} c_{\lambda} = 1$ , we get



$$1 - \text{Tr}[\rho_0 \Pi] \leq \sum_{\lambda, l, l'} c_\lambda |\rho_{ll'}^S| \frac{4 \|H_I\|^2}{\eta^2} \leq d_S \frac{4 \|H_I\|^2}{\eta^2}. \quad (\text{C10})$$

Choosing the truncation window with  $\eta = K\sqrt{8d_S}\|H_I\|$  and using the gentle measurement lemma leads to the main result

$$\|\rho_0 - \Pi\rho_0\Pi\|_1 \leq \frac{2}{K}. \quad (\text{C11})$$

It is easy to extend this to the weak distinguishability. Note that the trace distance is invariant under global unitaries; in particular, it is invariant under the Hamiltonian evolution. Therefore, expectation values of an observable  $A$  will be close since we have

$$\begin{aligned} |\text{Tr}[A\rho_t] - \text{Tr}[A\Pi\rho_t\Pi]| &\leq \|\rho_t - \Pi\rho_t\Pi\|_1 \|A\| \\ &= \|\rho_0 - \Pi\rho_0\Pi\|_1 \|A\| \\ &\leq \frac{2\|A\|}{K}, \end{aligned} \quad (\text{C12})$$

by using, in the first step, that for any two operators  $O$  and  $P$ , Holder's inequality implies  $\text{Tr}[OP] \leq \|OP\|_1 \leq \|O\|_1 \|P\|$ .

From the above results, we obtain

$$\tilde{D}_A(\rho_t, \Pi\rho_t\Pi) \leq \frac{1}{K^2}. \quad (\text{C13})$$

To conclude, it will be useful to relate the rank of  $\Pi$  to the rank of a projector  $P$  onto yet another extended subspace  $\mathcal{H}_{H_B+H_S}^{E_B, \Delta+2\|H_S\|+\eta+\eta'}$ . Since the rank of  $P$  is related to the spectrum of unperturbed system and bath Hamiltonians, it will prove simpler to calculate.

First, denoting by  $P^\perp$  the orthogonal projector to  $P$ , we get

$$\begin{aligned} \|\Pi - \Pi P\|_1 &= \|\Pi P^\perp\|_1 \\ &\leq \text{rank}(\Pi) \|\Pi P^\perp\| \\ &\leq \|\Pi\|_1 \frac{2\|H_I\|}{\eta'} = \frac{\|\Pi\|_1}{K}, \end{aligned} \quad (\text{C14})$$

by using the fact that for any operator  $\|Q\|_1 \leq \text{rank}(Q)\|Q\|$ , using Bhatia's theorem, and setting  $\eta' = 2K\|H_I\|$ . The triangle inequality then leads to

$$\begin{aligned} \|\Pi\|_1 &\leq \|\Pi P\|_1 + \|\Pi - \Pi P\|_1 \\ &\leq \|P\|_1 + \frac{\|\Pi\|_1}{K}. \end{aligned} \quad (\text{C15})$$

Recall that  $d_{\text{trunc}} = \text{rank}(\Pi) = \|\Pi\|_1$ . Hence,

$$d_{\text{trunc}} \leq \frac{\|P\|_1}{1 - \frac{1}{K}}. \quad (\text{C16})$$

Note that, since  $P$  projects onto the subspace  $\mathcal{H}_{H_B+H_S}^{E_B, \Delta+2\|H_S\|+\eta+\eta'}$  corresponding to a system that does not interact with the bath, we can denote the density of states of the bath by  $\nu_B(E)$ , and

$$\begin{aligned} \|P\|_1 &\leq d_S \int_{E_B - \frac{\Delta}{2} - \frac{\eta}{2} - \frac{\eta'}{2} - \|H_S\|}^{E_B + \frac{\Delta}{2} + \frac{\eta}{2} + \frac{\eta'}{2} + \|H_S\|} \nu_B(E) dE \\ &= d_S \int_{E_B - \frac{\Delta}{2} - (1 + \sqrt{2d_S})K\|H_I\| - \|H_S\|}^{E_B + \frac{\Delta}{2} + (1 + \sqrt{2d_S})K\|H_I\| + \|H_S\|} \nu_B(E) dE. \end{aligned} \quad (\text{C17})$$

The inequality comes from upper bounding the number of accessible states of the system by  $d_S$  and counting the states of the bath as if it could access all of the possible energies of the space  $\mathcal{H}_{H_B+H_S}^{E_B, \Delta+2\|H_S\|+\eta+\eta'}$ . The second line comes from using  $\eta' = 2K\|H_I\|$  and  $\eta = \sqrt{8d_S}K\|H_I\|$ .

## APPENDIX D: PROOF DETAILS OF THEOREM 8

We write the original state in the basis  $|j\rangle$  of eigenvectors of the full Hamiltonian  $H$  as

$$\rho_0 = \sum_{jk} \rho_{jk} |j\rangle\langle k|. \quad (\text{D1})$$

The truncated state is then

$$\Pi\rho_0\Pi = \sum_{jk \in J} \rho_{jk} |j\rangle\langle k|, \quad (\text{D2})$$

where  $\Pi$  projects to the truncated Hilbert space  $\mathcal{H}_{H_S+H_B+H_I}^{E_B, \Delta+2\|H_S\|+\eta}$ , and  $J$  is the set of eigenvalues of  $H_S + H_B + H_I$  restricted to such space.

We now expand the Hilbert space to include two new energy eigenvectors  $|j_{\min}\rangle$  and  $|j_{\max}\rangle$ , with corresponding energies  $E_{\min}$  and  $E_{\max}$ , respectively, such that the new Hamiltonian is

$$\bar{H} \equiv H + E_{\min}|j_{\min}\rangle\langle j_{\min}| + E_{\max}|j_{\max}\rangle\langle j_{\max}|. \quad (\text{D3})$$

Next, we define a new density matrix on the enlarged space by

$$\begin{aligned} \bar{\rho}_0 &\equiv \zeta \left( \Pi\rho_0\Pi + \frac{|x|}{2} |j_{\min}\rangle\langle j_{\min}| + \frac{x}{2} |j_{\min}\rangle\langle j_{\max}| \right. \\ &\quad \left. + \frac{x}{2} |j_{\max}\rangle\langle j_{\min}| + \frac{|x|}{2} |j_{\max}\rangle\langle j_{\max}| \right), \end{aligned} \quad (\text{D4})$$

where  $x$  is a real constant and  $\zeta$  is an appropriate normalization constant to ensure that  $\text{Tr}[\bar{\rho}_0] = 1$ . The above definition ensures that  $\bar{\rho}_0$  remains a positive operator. Note also that  $\Pi$  is orthogonal to  $|j_{\min}\rangle$  and  $|j_{\max}\rangle$  since the truncation we perform is in the original Hilbert space.

It will also be useful to define an observable  $\bar{A}$  in a similar way, as

$$\bar{A} \equiv \frac{1}{\zeta} \left( \Pi A \Pi + \frac{\|A\|}{2} |j_{\min}\rangle\langle j_{\min}| + \frac{\|A\|}{2} |j_{\min}\rangle\langle j_{\max}| + \frac{\|A\|}{2} |j_{\max}\rangle\langle j_{\min}| + \frac{\|A\|}{2} |j_{\max}\rangle\langle j_{\max}| \right). \quad (\text{D5})$$

In the derivation of Theorem 6, the commutators that appear in the denominator of Eq. (20) came from the standard deviation of the distribution  $p_\alpha$ , as explained in Sec. III. We intend to use  $\bar{\rho}_0$  and  $\bar{A}$  as the initial state and observable for our calculation, while proving that  $x$  can be taken such that

$$(i) \quad \bar{\sigma}_G^2 \geq \frac{1}{\bar{Q}\|\bar{A}\|} |\text{Tr}([\rho_0, H], H)A|,$$

$$(ii) \quad \{\bar{\rho}_0, \bar{A}, \bar{H}\} \text{ lead to approximately the same physics as } \{\rho_0, A, H\},$$

where  $\bar{Q}$  is the normalization factor for the distribution  $\bar{p}_\alpha$  corresponding to  $\bar{\rho}_0$  and  $\bar{A}$ .

Point (i) will lead to a result in Theorem 8 with commutators involving the original triple  $\{\rho_0, A, H\}$ , as desired, while point (ii) allows us to use the redefined state, observable, and Hamiltonian instead of the original ones.

### 1. Going from $\{\bar{\rho}_0, \bar{A}, \bar{H}\}$ to commutators involving $\{\rho_0, A, H\}$

Let us revisit Eq. (22), this time with the state  $\bar{\rho}_0$  and the observable  $\bar{A}$ . We can see

$$\begin{aligned} \bar{\sigma}_G^2 &= \sum_\alpha \bar{p}_\alpha G_\alpha^2 = \frac{1}{\bar{Q}\|\bar{A}\|} \sum_{j \neq k} |\bar{p}_{jk}| |\bar{A}_{kj}| (E_j - E_k)^2 \\ &= \frac{1}{\bar{Q}\|\bar{A}\|} \left( \sum_{jk \in J} |\rho_{jk}| |A_{kj}| (E_j - E_k)^2 + 2 \frac{|x| \|A\|}{2} G_{\max}^2 \right) \\ &\geq \frac{1}{\bar{Q}\|\bar{A}\|} \left( \left| \sum_{jk \in J} \rho_{jk} A_{kj} (E_j - E_k)^2 \right| + \frac{x \|A\|}{2} G_{\max}^2 \right), \quad (\text{D6}) \end{aligned}$$

where we defined the maximum energy gap  $G_{\max} = E_{\max} - E_{\min}$ .

We now *impose* that  $x$  is such that

$$\begin{aligned} &\frac{1}{\bar{Q}\|\bar{A}\|} \left( \left| \sum_{jk \in J} \rho_{jk} A_{kj} (E_j - E_k)^2 \right| + \frac{x \|A\|}{2} G_{\max}^2 \right) \\ &\equiv \frac{1}{\bar{Q}\|\bar{A}\|} \left| \sum_{jk} \rho_{jk} A_{kj} (E_j - E_k)^2 \right| \\ &= \frac{1}{\bar{Q}\|\bar{A}\|} |\text{Tr}([\rho_0, H], H)A|, \quad (\text{D7}) \end{aligned}$$

which together with Eq. (D6) already gives condition (i).

From the equation above, we obtain

$$\begin{aligned} x \frac{\|A\| G_{\max}^2}{2} &= \left| \sum_{jk} \rho_{jk} A_{kj} (E_j - E_k)^2 \right| \\ &\quad - \left| \sum_{jk \in J} \rho_{jk} A_{kj} (E_j - E_k)^2 \right| \\ &\leq \left| \sum_{jk} \rho_{jk} A_{kj} (E_j - E_k)^2 \right. \\ &\quad \left. - \sum_{jk \in J} \rho_{jk} A_{kj} (E_j - E_k)^2 \right| \\ &= |\text{Tr}([\rho_0 - \Pi \rho_0 \Pi, H], H)A| \\ &= |\text{Tr}([\rho_0 - \Pi \rho_0 \Pi, H'], H')A|, \quad (\text{D8}) \end{aligned}$$

where, in the last line, we define the auxiliary Hamiltonian  $H' \equiv H - [(E_{\max} + E_{\min})/2]1$ , shifted so that its spectrum is centered around 0. By noting that  $\|H'\| = \|H\| - [(E_{\max} + E_{\min})/2] = [(E_{\max} - E_{\min})/2] = [G_{\max}/2]$ , expanding the four terms in the above commutators, using the fact that for any two operators  $O$  and  $P$  one has  $\text{Tr}[OP] \leq \|O\| \|P\|_1$  from Holder's inequality, and using Eq. (C11), we end up with

$$\begin{aligned} x &\leq \frac{2}{\|A\| G_{\max}^2} 4 \|\rho_0 - \Pi \rho_0 \Pi\|_1 \|A\| \|H'\|^2 \\ &= 2 \|\rho_0 - \Pi \rho_0 \Pi\|_1 \leq \frac{4}{K}. \quad (\text{D9}) \end{aligned}$$

Similarly, we can show that  $-x \leq (4/K)$ , and hence,

$$|x| \leq \frac{4}{K}. \quad (\text{D10})$$

### 2. $\{\bar{\rho}_0, \bar{A}, \bar{H}\}$ give approximately the same physics as $\{\rho_0, A, H\}$

We now check that condition (ii) is also satisfied. Note that

$$\begin{aligned} &|\text{Tr}[\rho_0 A] - \text{Tr}[\bar{\rho}_0 \bar{A}]| \\ &\leq |\text{Tr}[\rho_0 A] - \text{Tr}[\Pi \rho_0 \Pi A]| + |\text{Tr}[\Pi \rho_0 \Pi A] - \text{Tr}[\bar{\rho}_0 \bar{A}]| \\ &= |\text{Tr}[\rho_0 A] - \text{Tr}[\Pi \rho_0 \Pi A]| + \left| \frac{x \|A\|}{2} + \frac{|x| \|A\|}{2} \right| \\ &\leq \|\rho_0 - \Pi \rho_0 \Pi\|_1 \|A\| + |x| \|A\| \\ &\leq \frac{2 \|A\|}{K} + \frac{4 \|A\|}{K}, \quad (\text{D11}) \end{aligned}$$

by using, in the third line,  $\text{Tr}[\bar{\rho}_0 \bar{A}] = \text{Tr}[A \Pi \rho_0 \Pi] + 2[(x \|A\|)/4] + 2[|x| \|A\|/4]$  (which comes from the definition of  $\bar{A}$  and  $\bar{\rho}_0$ ), and the fact that  $\Pi$  is orthogonal to  $|j_{\min}\rangle$  and  $|j_{\max}\rangle$ .

The above equation justifies the approach of defining the auxiliary state and observable since we proved that these mimic the original state and observable in the predictions.

The result can be translated into the weak distinguishability between  $\rho_t$  and  $\omega$ . By using the triangle inequality, a similar calculation as above, and the fact that the trace distance is invariant under unitary evolution, we see

$$\begin{aligned} |\text{Tr}[\rho_t A] - \text{Tr}[\omega A]|^2 &\leq |\text{Tr}[\rho_t A] - \text{Tr}[\bar{\rho}_t \bar{A}]|^2 \\ &\quad + |\text{Tr}[\bar{\rho}_t \bar{A}] - \text{Tr}[\bar{\omega} \bar{A}]|^2 \\ &\quad + |\text{Tr}[\bar{\omega} \bar{A}] - \text{Tr}[\omega A]|^2 \\ &\leq |\text{Tr}[\bar{\rho}_t \bar{A}] - \text{Tr}[\bar{\omega} \bar{A}]|^2 + 2 \left( \frac{6\|A\|}{K} \right)^2, \end{aligned} \quad (\text{D12})$$

where  $\bar{\rho}_t = e^{-i\bar{H}t} \rho_t e^{i\bar{H}t}$  and  $\bar{\omega}$  is the corresponding dephased state. This implies

$$\tilde{\mathcal{D}}_A(\rho_t, \omega) \leq \frac{\|\bar{A}\|^2}{\|A\|^2} \tilde{\mathcal{D}}_A(\bar{\rho}_t, \bar{\omega}) + \frac{18}{K^2}. \quad (\text{D13})$$

For the term  $\tilde{\mathcal{D}}_A(\bar{\rho}_t, \bar{\omega})$ , we can apply Propositions 4 and 5, and condition (i) to get

$$\begin{aligned} \langle \tilde{\mathcal{D}}_A(\bar{\rho}_t, \bar{\omega}) \rangle_T &\leq \frac{\pi a(\epsilon) \bar{\mathcal{Q}}^2}{T \bar{\sigma}_G} + \pi \delta(\epsilon) \bar{\mathcal{Q}}^2 \\ &\leq \frac{\pi \|\bar{A}\|^{1/2} a(\epsilon) \bar{\mathcal{Q}}^{5/2}}{T \sqrt{|\text{Tr}([\rho_0, H], H)A|}} + \pi \delta(\epsilon) \bar{\mathcal{Q}}^2. \end{aligned} \quad (\text{D14})$$

The last inequality comes from our convenient construction of  $\bar{\rho}_0$  and  $\bar{A}$ , which is specifically designed for this.

### 3. The factor $\bar{\mathcal{Q}}$ for $\bar{\rho}_0$ and $\bar{A}$

It is easy to see that the factor  $\bar{\mathcal{Q}}$  for the auxiliary state and observable satisfies

$$\begin{aligned} \bar{\mathcal{Q}} &\equiv \sum_{j \neq k} |\bar{\rho}_{jk}| \frac{|\bar{A}_{kj}|}{\|\bar{A}\|} \\ &= \sum_{jk \in J} |\rho_{jk}| \frac{|A_{kj}|}{\|A\|} + 2 \frac{|x|}{2} \frac{\|A\|}{2\|\bar{A}\|} \\ &= \frac{\|A\|}{\|\bar{A}\|} \left( Q_{\text{trunc}} + \frac{|x|}{2} \right) \\ &\leq \frac{\|A\|}{\|\bar{A}\|} \left( Q_{\text{trunc}} + \frac{2}{K} \right), \end{aligned} \quad (\text{D15})$$

where  $Q_{\text{trunc}}$  is the normalization constant of the distribution that results from  $\Pi \rho_0 \Pi$  and  $A$ .

The above bound, plus Eqs. (D13) and (D14), results in

$$\begin{aligned} \langle \tilde{\mathcal{D}}_A(\rho_t, \omega) \rangle_T &\leq \frac{\|\bar{A}\|^2}{\|A\|^2} \frac{\pi a(\epsilon) \|\bar{A}\|^{1/2} \bar{\mathcal{Q}}^{5/2}}{T \sqrt{|\text{Tr}([\rho_0, H], H)A|}} \\ &\quad + \frac{\|\bar{A}\|^2}{\|A\|^2} \pi \delta(\epsilon) \bar{\mathcal{Q}}^2 + \frac{18}{K^2} \\ &\leq \frac{\pi a(\epsilon) \|A\|^{1/2} (Q_{\text{trunc}} + \frac{2}{K})^{5/2}}{T \sqrt{|\text{Tr}([\rho_0, H], H)A|}} \\ &\quad + \pi \delta(\epsilon) \left( Q_{\text{trunc}} + \frac{2}{K} \right)^2 + \frac{18}{K^2}, \end{aligned} \quad (\text{D16})$$

which, defining  $Q_2 = Q_{\text{trunc}} + (2/K)$  to simplify notation, gives the first part of Theorem 8.

In order to finish the Theorem's proof, we upper bound  $Q_{\text{trunc}}$ . From Eq. (23), we see that for this state

$$\begin{aligned} Q_{\text{trunc}}^2 &\leq \text{Tr}[(\Pi \rho_0 \Pi)^2] \text{Tr}[\Pi] \\ &= \text{Tr}[\Pi \rho_0 \Pi \rho_0] d_{\text{trunc}} \\ &\leq \sqrt{\text{Tr}[\Pi \rho_0^2 \Pi] \text{Tr}[\rho_0 \Pi^2 \rho_0]} d_{\text{trunc}} \\ &\leq \|\Pi\| \text{Tr}[\rho_0^2] d_{\text{trunc}} \\ &= \text{Tr}[\rho_0^2] d_{\text{trunc}}, \end{aligned} \quad (\text{D17})$$

by using the definition of  $d_{\text{trunc}}$ , the Cauchy-Schwarz inequality, and the fact that, for two positive semidefinite matrices,  $\text{Tr}[OP] \leq \|O\| \text{Tr}[P]$ .

Since  $\rho_0 = \rho_S \otimes [(\frac{1}{d_B}) / (d_B^A)]$ , we see that

$$\text{Tr}[\rho_0^2] = \frac{\text{Tr}_S[\rho_S^2]}{d_B^A}. \quad (\text{D18})$$

In the main text, we found [Eq. (37)]

$$d_{\text{trunc}} \leq \frac{d_S d_B^A}{(1 - \frac{1}{K})(1 - e^{-\beta \Delta})} e^{\beta \|H_S\| + (1 + \sqrt{2d_S}) K \beta \|H_I\|}, \quad (\text{D19})$$

which leads to

$$\begin{aligned} Q_{\text{trunc}}^2 &\leq \text{Tr}[\rho_0^2] d_{\text{trunc}} \\ &\leq \frac{d_S \text{Tr}_S[\rho_S^2]}{(1 - \frac{1}{K})(1 - e^{-\beta \Delta})} e^{\beta \|H_S\| + (1 + \sqrt{2d_S}) K \beta \|H_I\|}. \end{aligned} \quad (\text{D20})$$

Substituting this back into the definition for  $Q_2$  gives Eq. (40).

### APPENDIX E: TYPICAL BEHAVIOR FOR ENVIRONMENT IN A PURE INITIAL STATE—PROPOSITION 9

With the initial state

$$\rho_0^U = \rho_S \otimes U|\psi\rangle\langle\psi|U^\dagger = \rho_S \otimes \rho_B^U, \quad (\text{E1})$$

we focus on the average over all unitaries  $U$  within the subspace  $\mathcal{H}_{H_B}^{E_B, \Delta}$  of  $\tilde{D}_A(\rho_t^U, \omega^U)$ :

$$\langle \tilde{D}_A(\rho_t^U, \omega^U) \rangle_U = \frac{\langle |\text{Tr}[\rho_t^U A] - \text{Tr}[\omega^U A]|^2 \rangle_U}{4\|A\|^2}. \quad (\text{E2})$$

By shifting the time dependencies to the observable  $A$ , we get

$$\text{Tr}[\rho_t^U A] = \text{Tr}[\rho_0^U e^{iHt} A e^{-iHt}] \equiv \text{Tr}[\rho_0^U A(t)]. \quad (\text{E3})$$

If  $A_{\text{eq}}$  is the infinite-time-averaged  $A(t)$ , we have  $\text{Tr}[\omega^U A] = \text{Tr}[\rho_0^U A_{\text{eq}}]$ . Then,

$$\begin{aligned} \langle \tilde{D}_A(\rho_t^U, \omega^U) \rangle_U &= \frac{1}{4\|A\|^2} \langle |\text{Tr}[\rho_0^U (A(t) - A_{\text{eq}})]|^2 \rangle_U \\ &= \frac{1}{4\|A\|^2} \langle |\text{Tr}_B[\rho_B^U \text{Tr}_S[\rho_S \otimes \mathbb{1}_B (A(t) - A_{\text{eq}})]]|^2 \rangle_U \\ &= \langle |\text{Tr}_B[\rho_B^U C]|^2 \rangle_U \\ &= \text{Tr}_{B^{\otimes 2}}[\langle \rho_B^U \otimes \rho_B^U \rangle_U C \otimes C], \end{aligned} \quad (\text{E4})$$

where, in the third line, we have used  $\rho_B^U = \mathbb{1}_B^{\Delta} \rho_B^U \mathbb{1}_B^{\Delta}$  to write  $C \equiv [1/(2\|A\|)] \text{Tr}_S[\rho_S \otimes \mathbb{1}_B^{\Delta} (A(t) - A_{\text{eq}}) \mathbb{1}_S \otimes \mathbb{1}_B^{\Delta}]$  as an observable acting on the microcanonical window of the bath Hilbert space.

Via the same calculations as used in the Appendix of Ref. [3], we get

$$\begin{aligned} \langle \rho_B^U \otimes \rho_B^U \rangle_U &= \langle U^{\otimes 2} (|\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi|) (U^{\otimes 2})^\dagger \rangle_U \\ &= \alpha \Pi_s + \beta \Pi_a, \end{aligned} \quad (\text{E5})$$

where  $\Pi_s = \{[(\mathbb{1}_B^{\Delta})^{\otimes 2} + \$]/2\}$  and  $\Pi_a = \{[(\mathbb{1}_B^{\Delta})^{\otimes 2} - \$]/2\}$  project onto the symmetric and antisymmetric subspaces, and  $\$$  is the swap operator on  $\mathcal{H}_{H_B}^{E_B, \Delta} \otimes \mathcal{H}_{H_B}^{E_B, \Delta}$ , defined by  $\$|\phi_1\rangle|\phi_2\rangle = |\phi_2\rangle|\phi_1\rangle$ . Since

$$\begin{aligned} \text{Tr}[\langle \rho_B^U \otimes \rho_B^U \rangle_U \Pi_a] &= \frac{1}{2} \langle (1 - \text{Tr}[\rho_B^U \otimes \rho_B^U \$]) \rangle_U \\ &= \frac{1}{2} \langle (1 - \text{Tr}[(\rho_B^U)^2]) \rangle_U \\ &= 0, \end{aligned} \quad (\text{E6})$$

we see that  $\beta = 0$ , and from  $\text{Tr}[\langle \rho_B^U \otimes \rho_B^U \rangle_U] = 1$ , we obtain  $\alpha = \{2/[d_B^{\Delta}(d_B^{\Delta} + 1)]\}$ , which leads to the simple expression  $\langle \rho_0^U \otimes \rho_0^U \rangle_U = \{2/[d_B^{\Delta}(d_B^{\Delta} + 1)]\} \Pi_s$ . Then,

$$\begin{aligned} \langle \tilde{D}_A(\rho_t^U, \omega^U) \rangle_U &= \frac{2}{d_B^{\Delta}(d_B^{\Delta} + 1)} \text{Tr}_{B^{\otimes 2}}[\Pi_s C \otimes C] \\ &= \frac{1}{d_B^{\Delta}(d_B^{\Delta} + 1)} (\text{Tr}_{B^{\otimes 2}}[(\mathbb{1}_B^{\Delta})^{\otimes 2} C \otimes C] \\ &\quad + \text{Tr}_{B^{\otimes 2}}[\$ C \otimes C]) \\ &= \frac{1}{d_B^{\Delta}(d_B^{\Delta} + 1)} (\text{Tr}_B[C]^2 + \text{Tr}_B[C^2]). \end{aligned} \quad (\text{E7})$$

The operator  $C$  is of the form  $C = \text{Tr}_S[O]$ , with  $O$  Hermitian and acting only on the microcanonical window, from its definition above. Any such operator can be written as

$$O = \sum_{j=0}^{d_S^2-1} \sum_{k=0}^{(d_B^{\Delta})^2-1} a_{jk} X_j Y_k, \quad (\text{E8})$$

where  $a_{jk}$  are real coefficients, and  $\{X_j\}$  and  $\{Y_k\}$  are orthonormal bases of Hermitian operators on the system and microcanonical window, respectively [50]. They satisfy

$$\begin{aligned} \text{Tr}_S[X_j X_{j'}] &= \delta_{jj'}, \quad \forall \{j, j'\} = (0, \dots, d_S^2 - 1), \\ \text{Tr}_B[Y_k Y_{k'}] &= \delta_{kk'}, \quad \forall \{k, k'\} = (0, \dots, (d_B^{\Delta})^2 - 1), \end{aligned} \quad (\text{E9})$$

with  $X_0 = [(\mathbb{1}_S)/\sqrt{d_S}]$  and  $Y_0 = [(\mathbb{1}_B^{\Delta})/\sqrt{d_B^{\Delta}}]$ , while all other operators have trace 0. With these definitions, we can write

$$\begin{aligned} \text{Tr}_B[C^2] &= \text{Tr}_B[\text{Tr}_S[O] \text{Tr}_S[O]] \\ &= \sum_{jj'=0}^{d_S^2-1} \sum_{kk'=0}^{(d_B^{\Delta})^2-1} a_{jk} a_{j'k'} \text{Tr}_B[\text{Tr}_S[X_j] \text{Tr}_S[X_{j'}] Y_k Y_{k'}] \\ &= \sum_{kk'=0}^{(d_B^{\Delta})^2-1} a_{0k} a_{0k'} \text{Tr}_B[Y_k Y_{k'}] \text{Tr}_S[X_0] \text{Tr}_S[X_0] \\ &= d_S \sum_{kk'=0}^{(d_B^{\Delta})^2-1} a_{0k} a_{0k'} \delta_{kk'} = d_S \sum_{k=0}^{(d_B^{\Delta})^2-1} a_{0k}^2 \\ &\leq d_S \sum_{j=0}^{d_S^2-1} \sum_{k=0}^{(d_B^{\Delta})^2-1} a_{jk}^2 = d_S \text{Tr}[O^2]. \end{aligned} \quad (\text{E10})$$

From this result, and the definition of  $C$  above, we obtain



$$\begin{aligned}
\text{Tr}_B[C^2] &\leq \frac{d_S}{4\|A\|^2} \text{Tr}[(\rho_S \otimes \mathbb{1}_B^\Delta(A(t) - A_{\text{eq}}))^2] \\
&\leq \frac{d_S}{4\|A\|^2} \sqrt{\text{Tr}[(\rho_S \otimes \mathbb{1}_B^\Delta)(A(t) - A_{\text{eq}})^2(\rho_S \otimes \mathbb{1}_B^\Delta)]} \\
&\quad \times \sqrt{\text{Tr}[(A(t) - A_{\text{eq}})(\rho_S \otimes \mathbb{1}_B^\Delta)(A(t) - A_{\text{eq}})]} \\
&\leq \frac{d_S}{4\|A\|^2} \text{Tr}[(\rho_S \otimes \mathbb{1}_B^\Delta)^2] \|A(t) - A_{\text{eq}}\|^2 \\
&\leq d_S d_B^\Delta
\end{aligned} \tag{E11}$$

by using Cauchy-Schwarz inequality on line 2, the fact that for positive semidefinite operators  $\text{Tr}[PQ] \leq \text{Tr}[P]\|Q\|$  on line 3, and the triangle inequality plus  $\|A_{\text{eq}}\| \leq \|A(t)\| = \|A\|$  on the last line.

With this result, we see that Eq. (E7) becomes

$$\begin{aligned}
\langle \tilde{D}_A(\rho_t^U, \omega^U) \rangle_U &= \frac{1}{d_B^\Delta(d_B^\Delta + 1)} (\text{Tr}_B[C]^2 + \text{Tr}_B[C^2]) \\
&\leq \frac{1}{d_B^\Delta(d_B^\Delta + 1)} \left( \frac{1}{4\|A\|^2} \text{Tr}[\rho_S \otimes \mathbb{1}_B^\Delta(A(t) - A_{\text{eq}})]^2 + d_S d_B^\Delta \right) \\
&= \frac{1}{d_B^\Delta(d_B^\Delta + 1)} \left( \frac{(d_B^\Delta)^2}{4\|A\|^2} \text{Tr}[\rho_S \otimes \frac{\mathbb{1}_B^\Delta}{d_B^\Delta}(A(t) - A_{\text{eq}})]^2 + d_S d_B^\Delta \right) \\
&= \frac{d_B^\Delta}{d_B^\Delta + 1} \frac{\text{Tr}[\rho_0(A(t) - A_{\text{eq}})]^2}{4\|A\|^2} + \frac{d_S}{d_B^\Delta + 1},
\end{aligned} \tag{E12}$$

where  $\rho_0 \equiv \rho_S \otimes [(\mathbb{1}_B^\Delta)/d_B^\Delta]$ . We can now shift the time dependence back to the state to get

$$\begin{aligned}
\langle \tilde{D}_A(\rho_t^U, \omega^U) \rangle_U &\leq \frac{d_B^\Delta}{d_B^\Delta + 1} \frac{\text{Tr}[(\rho_t - \omega)A]^2}{4\|A\|^2} + \frac{d_S}{d_B^\Delta + 1} \\
&\leq \frac{\text{Tr}[(\rho_t - \omega)A]^2}{4\|A\|^2} + \frac{d_S}{d_B^\Delta} \\
&= \tilde{D}_A(\rho_t, \omega) + \frac{d_S}{d_B^\Delta},
\end{aligned} \tag{E13}$$

which proves our claim. The bound  $[1/(d_B^\Delta + 1)] \leq [1/(d_B^\Delta)]$  is only for presentation reasons and does not change the result much since  $d_B^\Delta \gg d_S > 1$  in the regime we are interested in.

## APPENDIX F: DISTRIBUTION $p_\alpha$ FOR A SPIN RING

In order to get a deeper grasp of the behavior of  $p_\alpha$ , we simulated  $L$  interacting spin  $1/2$ 's, with a Hamiltonian given by

$$H = \Omega \sum_{\lambda=1}^L \sigma_\lambda^z + \gamma \Omega \sum_{\lambda=1}^L \sigma_\lambda^x \otimes \sigma_{\lambda+1}^x, \tag{F1}$$

where  $\sigma_\lambda^z$  and  $\sigma_\lambda^x$  are the Pauli  $z$  and  $x$  operators for the spin  $\lambda$ , and we adopt the notation  $\sigma_{L+1}^x = \sigma_1^x$ . The spin  $\lambda = 1$  is taken to represent the system  $\mathcal{S}$ , and we focus on an observable  $A_x$  and initial state  $\rho$  given by

$$A_x = \sigma_1^x \otimes \bigotimes_{\lambda=2}^L \mathbb{1}_\lambda, \quad \rho_0 = |1\rangle\langle 1| \otimes \bigotimes_{\lambda=2}^L \frac{\mathbb{1}_\lambda}{2}, \tag{F2}$$

the latter representing a bath in a maximally mixed initial state and the system in the eigenvector  $|1\rangle$  of  $\sigma_1^z$  with eigenvalue 1.

Figure 3 depicts the normalized distribution  $p_\alpha = (1/Q)[(|\rho_{jk} A_{kj}|)/\|A_x\|]$  as a function of the energy gaps  $G_\alpha = (E_j - E_k)$  as the number of spins increases, illustrating the transition between a distribution with several distinct peaks and a unimodal distribution. In Fig. 4, we plot  $a(\epsilon)$  and  $\delta(\epsilon)$  for different values of  $L$ , illustrating their decrease with increasing  $L$ , for most values of  $\epsilon$ . Thus, even for moderate sizes of the bath, one can find an interval  $\epsilon$  such that  $\delta(\epsilon) \ll 1$  and  $a \sim 1$ . The first condition is necessary for Theorem 1 and the subsequent results in order to imply that equilibration occurs, while the second condition is necessary to ensure that the equilibration time scale does not grow for increasing bath sizes (see discussion after Theorem 8).

For example, in this model, we find that for  $L = 9$ , one can take  $\epsilon$  such that  $a(\epsilon) = 1$  and  $\delta(\epsilon) \approx 0.006$ . Then, Theorem 1 gives

$$\langle \tilde{D}_A(\rho_t, \omega) \rangle_T \leq \frac{2^{5/4} \pi \|A_S\|^{1/2}}{T \sqrt{|\text{Tr}([\rho_0, H_S + H_I], H_S + H_I) A_S|}} + 0.04, \tag{F3}$$

with an upper bound on the equilibration time scale

$$T_{\text{eq}} \equiv \frac{2^{5/4} \pi \|A_S\|^{1/2}}{\sqrt{|\text{Tr}([\rho_0, H_S + H_I], H_S + H_I) A_S|}}, \tag{F4}$$

dictated by the observable, initial state, and system and interaction Hamiltonians, which are straightforward to calculate.

Figure 5 shows similar behavior of  $a(\epsilon)$  and  $\delta(\epsilon)$  for a spin ring with random couplings. We simulated a Hamiltonian

$$H = \Omega \sum_{\lambda=1}^L \sigma_\lambda^z + \Omega \sum_{\lambda=1}^L K_\lambda \sigma_\lambda^x \otimes \sigma_{\lambda+1}^x, \tag{F5}$$

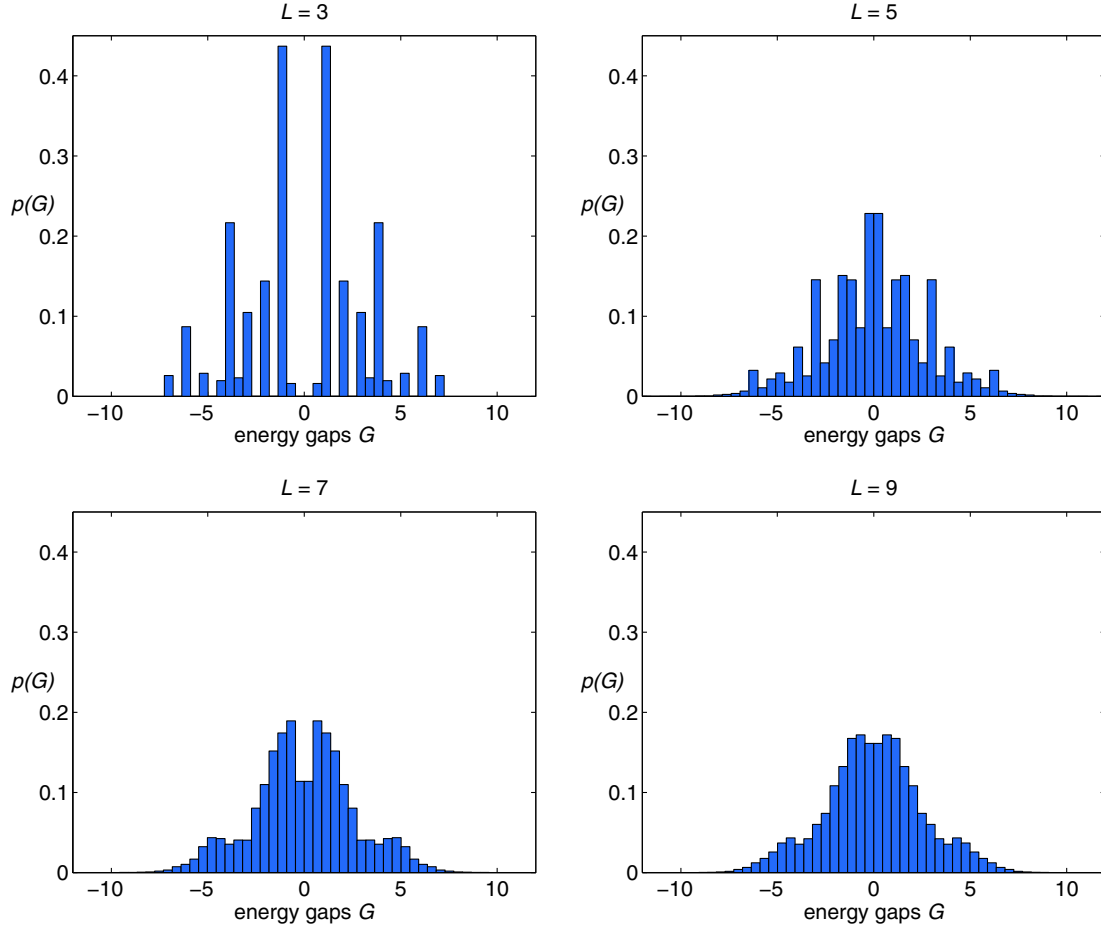


FIG. 3. Normalized histogram of the distribution  $p_\alpha$  for an observable  $A_x = \sigma_1^x \otimes \bigotimes_{\lambda=2}^L \mathbb{1}_\lambda$  and initial state  $\rho_0 = |1\rangle\langle 1| \otimes \bigotimes_{\lambda=2}^L \mathbb{1}_\lambda/2$  as a function of the energy gaps for a spin ring with coupling strength  $\gamma = 1.1\Omega$ , for an increasing number  $L$  of spins. For small  $L$ , the distribution  $p_\alpha$  is composed of distinct peaks. On the other hand,  $p_\alpha$  is spread over more values as the size of the system increases. At the same time, as  $L$  increases, the distribution becomes more distinctly unimodal.

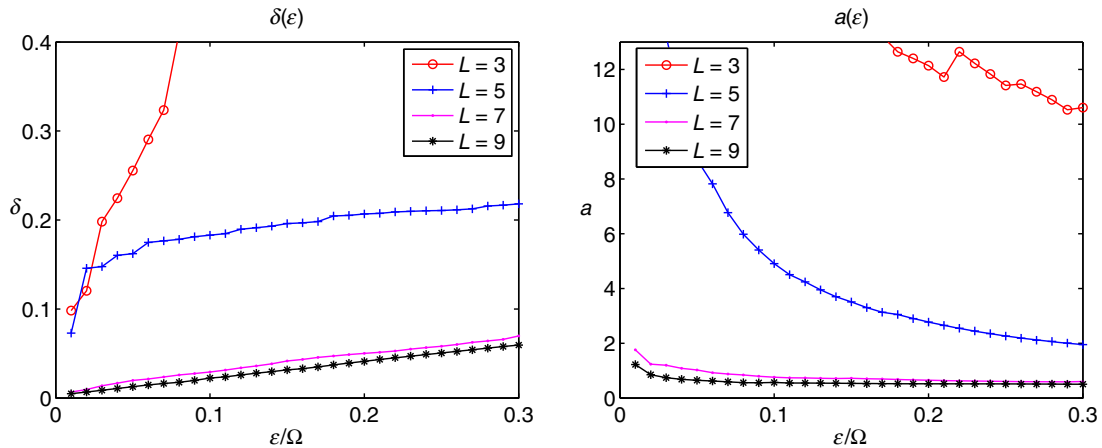


FIG. 4. Plots of  $\delta(\epsilon)$  and  $a(\epsilon)$ , defined in Proposition 5, for an observable  $A_x = \sigma_1^x \otimes \bigotimes_{\lambda=2}^L \mathbb{1}_\lambda$  and initial state  $\rho_0 = |1\rangle\langle 1| \otimes \bigotimes_{\lambda=2}^L \mathbb{1}_\lambda/2$ , for a spin ring with coupling constant  $\gamma = 1.1\Omega$ . There is a trade-off between having small  $\delta(\epsilon)$  and small  $a(\epsilon)$ . However, for a fixed energy gap interval  $\epsilon$ , both  $a(\epsilon)$  and  $\delta(\epsilon)$  decrease as  $L$  increases. Hence, as  $L$  increases, it becomes possible to find  $\epsilon$  such that  $a \sim 1$  and  $\delta \ll 1$ . Indeed, for  $L = \{3, 5, 7, 9\}$ , we have  $a(\epsilon) \approx 1$  and  $\delta(\epsilon) \approx \{1, 0.62, 0.02, 0.006\}$  for  $\epsilon \approx \{3.26, 1.65, 0.05, 0.02\}\Omega$ , respectively.

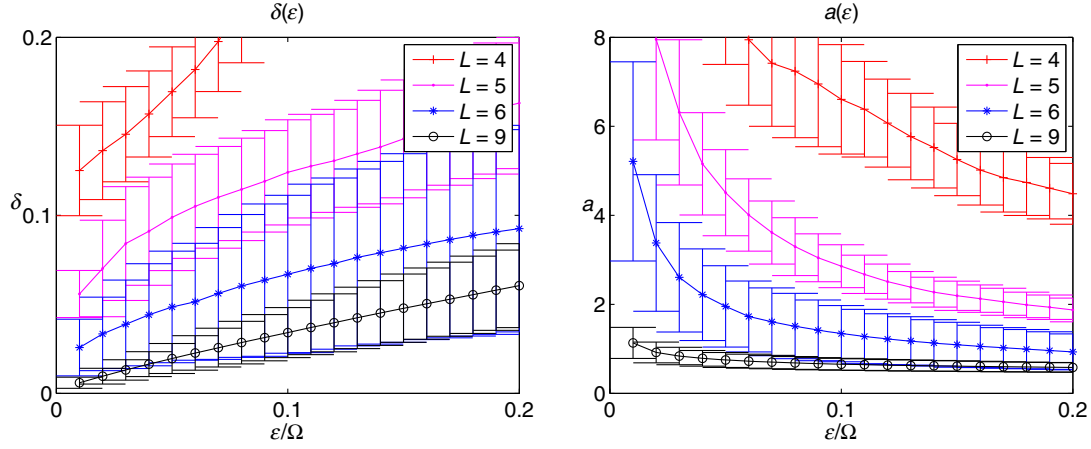


FIG. 5. Plots of average  $\delta(\epsilon)$  and average  $a(\epsilon)$  and their respective standard deviations (after 1000 realizations) for an observable  $A_z = \sigma_1^z \otimes \bigotimes_{\lambda=2}^L \mathbb{1}_\lambda$  and initial state  $\rho_0 = |1\rangle\langle 1| \otimes \bigotimes_{\lambda=2}^L \mathbb{1}_\lambda/2$ , for a spin ring with random couplings  $K_\lambda$  with mean  $\gamma = \Omega$  and standard deviation  $w = 0.2 \Omega$ . Once again, both  $a(\epsilon)$  and  $\delta(\epsilon)$  decrease as  $L$  increases, for a fixed energy gap interval  $\epsilon$ . Hence, as  $L$  increases, it becomes possible to find  $\epsilon$  such that  $a \sim 1$  and  $\delta \ll 1$ .

where the couplings  $K_\lambda$  are drawn at random from a Gaussian distribution with mean  $\gamma$  and standard deviation  $w$ . For completeness, in Fig. 5 we focus on an observable  $A_z$  and the same initial state as above:

$$A_z = \sigma_1^z \otimes \bigotimes_{\lambda=2}^L \mathbb{1}_\lambda, \rho_0 = |1\rangle\langle 1| \otimes \bigotimes_{\lambda=2}^L \frac{\mathbb{1}_\lambda}{2}. \quad (\text{F6})$$

- [1] A.J. Short and T.C. Farrelly, *Quantum Equilibration in Finite Time*, *New J. Phys.* **14**, 013063 (2012).
- [2] S. Goldstein, T. Hara, and H. Tasaki, *Time Scales in the Approach to Equilibrium of Macroscopic Quantum Systems*, *Phys. Rev. Lett.* **111**, 140401 (2013).
- [3] A. S. L. Malabarba, L. P. García-Pintos, N. Linden, T. C. Farrelly, and A. J. Short, *Quantum Systems Equilibrate Rapidly for Most Observables*, *Phys. Rev. E* **90**, 012121 (2014).
- [4] S. Goldstein, T. Hara, and H. Tasaki, *Extremely Quick Thermalization in a Macroscopic Quantum System for a Typical Nonequilibrium Subspace*, *New J. Phys.* **17**, 045002 (2015).
- [5] P. Reimann, *Typical Fast Thermalization Processes in Closed Many-Body Systems*, *Nat. Commun.* **7**, 10821 (2016).
- [6] M. Cramer, C. M. Dawson, J. Eisert, and T. J. Osborne, *Exact Relaxation in a Class of Nonequilibrium Quantum Lattice Systems*, *Phys. Rev. Lett.* **100**, 030602 (2008).
- [7] M. Diez, N. Chancellor, S. Haas, L. C. Venuti, and P. Zanardi, *Local Quenches in Frustrated Quantum Spin Chains: Global Versus Subsystem Equilibration*, *Phys. Rev. A* **82**, 032113 (2010).
- [8] M. Kastner, *Diverging Equilibration Times in Long-Range Quantum Spin Models*, *Phys. Rev. Lett.* **106**, 130601 (2011).
- [9] Vinayak and M. Žnidarič, *Subsystem Dynamics under Random Hamiltonian Evolution*, *J. Phys. A* **45**, 125204 (2012).
- [10] F. G. S. L. Brandão, P. Ćwikliński, M. Horodecki, P. Horodecki, J. K. Korbicz, and M. Mozrymas, *Convergence to Equilibrium under a Random Hamiltonian*, *Phys. Rev. E* **86**, 031101 (2012).
- [11] L. Masanes, A. J. Roncaglia, and A. Acín, *Complexity of Energy Eigenstates as a Mechanism for Equilibration*, *Phys. Rev. E* **87**, 032137 (2013).
- [12] M. van den Worm, B. C. Sawyer, J. J. Bollinger, and M. Kastner, *Relaxation Timescales and Decay of Correlations in a Long-Range Interacting Quantum Simulator*, *New J. Phys.* **15**, 083007 (2013).
- [13] L. Campos Venuti, S. Yeshwanth, and S. Haas, *Equilibration Times in Clean and Noisy Systems*, *Phys. Rev. A* **87**, 032108 (2013).
- [14] M. Fuchs, J. Schliemann, and B. Trauzettel, *Ultralong Spin Decoherence Times in Graphene Quantum Dots with a Small Number of Nuclear Spins*, *Phys. Rev. B* **88**, 245441 (2013).
- [15] P. R. Zangara, A. D. Dente, E. J. Torres-Herrera, H. M. Pastawski, A. Iucci, and L. F. Santos, *Time Fluctuations in Isolated Quantum Systems of Interacting Particles*, *Phys. Rev. E* **88**, 032913 (2013).
- [16] E. J. Torres-Herrera and L. F. Santos, *Quench Dynamics of Isolated Many-Body Quantum Systems*, *Phys. Rev. A* **89**, 043620 (2014).
- [17] L. Campos Venuti and P. Zanardi, *Theory of Temporal Fluctuations in Isolated Quantum Systems*, *Int. J. Mod. Phys. B* **29**, 1530008 (2015).
- [18] Hyungwon Kim, Mari Carmen Bañuls, J. Ignacio Cirac, Matthew B. Hastings, and David A. Huse, *Slowest Local Operators in Quantum Spin Chains*, *Phys. Rev. E* **92**, 012128 (2015).
- [19] M. Kastner and M. van den Worm, *Relaxation Timescales and Prethermalization in D-Dimensional Long-Range Quantum Spin Models*, *Phys. Scr.* **T165**, 014039 (2015).

- [20] C. Gogolin and J. Eisert, *Equilibration, Thermalisation, and the Emergence of Statistical Mechanics in Closed Quantum Systems*, *Rep. Prog. Phys.* **79**, 056001 (2016).
- [21] J. Eisert, M. Friesdorf, and C. Gogolin, *Quantum Many-Body Systems out of Equilibrium*, *Nat. Phys.* **11**, 124 (2015).
- [22] J. Goold, M. Huber, A. Riera, L. del Rio, and P. Skrzypczyk, *The Role of Quantum Information in Thermodynamics—A Topical Review*, *J. Phys. A* **49**, 143001 (2016).
- [23] The calculations are greatly simplified in the microcanonical ensemble, although the results also likely hold in the canonical ensemble since the state of the bath is highly mixed for not-too-low temperatures in this case, too.
- [24] Notice that this is a more general process than approaching thermal equilibrium since we do not restrict the equilibrium state to being thermal. Moreover, we do not focus on the particular way in which the system approaches equilibrium, and, in particular, the time decay to equilibrium need not be exponential.
- [25] N. Linden, S. Popescu, A. J. Short, and A. Winter, *Quantum Mechanical Evolution Towards Thermal Equilibrium*, *Phys. Rev. E* **79**, 061103 (2009).
- [26] S. Trotzky, Y.-A. Chen, A. Flesch, I. P. McCulloch, U. Schollwöck, J. Eisert, and I. Bloch, *Probing the Relaxation Towards Equilibrium in an Isolated Strongly Correlated One-Dimensional Bose Gas*, *Nat. Phys.* **8**, 325 (2012).
- [27] T. Langen, R. Geiger, M. Kuhnert, B. Rauer, and J. Schmiedmayer, *Local Emergence of Thermal Correlations in an Isolated Quantum Many-Body System*, *Nat. Phys.* **9**, 640 (2013).
- [28] R. Geiger, T. Langen, I. E. Mazets, and J. Schmiedmayer, *Local Relaxation and Light-Cone-Like Propagation of Correlations in a Trapped One-Dimensional Bose Gas*, *New J. Phys.* **16**, 053034 (2014).
- [29] P. Richerme, Z.-X. Gong, A. Lee, C. Senko, J. Smith, M. Foss-Feig, S. Michalakakis, A. V. Gorshkov, and C. Monroe, *Non-local Propagation of Correlations in Quantum Systems with Long-Range Interactions*, *Nature (London)* **511**, 198 (2014).
- [30] P. Jurcevic, B. P. Lanyon, P. Hauke, C. Hempel, P. Zoller, R. Blatt, and C. F. Roos, *Quasiparticle Engineering and Entanglement Propagation in a Quantum Many-Body System*, *Nature (London)* **511**, 202 (2014).
- [31] P. Reimann, *Foundation of Statistical Mechanics under Experimentally Realistic Conditions*, *Phys. Rev. Lett.* **101**, 190403 (2008).
- [32] The spectral norm is the maximum singular value,  $\|A\| = \max_{|x| \neq 0} [(|Ax|)/|x|]$ , with  $|x|$  the Euclidean norm for vectors. We will also make use of the trace norm, defined by  $\|A\|_1 = \text{Tr}[\sqrt{AA^\dagger}]$ , and the Hilbert-Schmidt norm  $\|A\|_2 = \sqrt{\text{Tr}[AA^\dagger]}$ .
- [33] J. M. Deutsch, *Quantum Statistical Mechanics in a Closed System*, *Phys. Rev. A* **43**, 2046 (1991).
- [34] M. Srednicki, *The Approach to Thermal Equilibrium in Quantized Chaotic Systems*, *J. Phys. A* **32**, 1163 (1999).
- [35] M. Rigol, V. Dunjko, V. Yurovsky, and M. Olshanii, *Relaxation in a Completely Integrable Many-Body Quantum System: An Ab Initio Study of the Dynamics of the Highly Excited States of 1D Lattice Hard-Core Bosons*, *Phys. Rev. Lett.* **98**, 050405 (2007).
- [36] M. Rigol, V. Dunjko, and M. Olshanii, *Thermalization and Its Mechanism for Generic Isolated Quantum Systems*, *Nature (London)* **452**, 854 (2008).
- [37] M. Rigol, *Breakdown of Thermalization in Finite One-Dimensional Systems*, *Phys. Rev. Lett.* **103**, 100403 (2009).
- [38] M. Rigol, *Dynamics and Thermalization in Correlated One-Dimensional Lattice Systems*, *arXiv:1008.1930*.
- [39] W. Beugeling, R. Moessner, and Masudul Haque, *Finite-Size Scaling of Eigenstate Thermalization*, *Phys. Rev. E* **89**, 042112 (2014).
- [40] R. Steinigeweg, A. Khodja, H. Niemeyer, C. Gogolin, and J. Gemmer, *Pushing the Limits of the Eigenstate Thermalization Hypothesis towards Mesoscopic Quantum Systems*, *Phys. Rev. Lett.* **112**, 130403 (2014).
- [41] A. Khodja, R. Steinigeweg, and J. Gemmer, *Relevance of the Eigenstate Thermalization Hypothesis for Thermal Relaxation*, *Phys. Rev. E* **91**, 012120 (2015).
- [42] H. Kim, T. N. Ikeda, and D. A. Huse, *Testing Whether All Eigenstates Obey the Eigenstate Thermalization Hypothesis*, *Phys. Rev. E* **90**, 052105 (2014).
- [43] M. Feingold and A. Peres, *Distribution of Matrix Elements of Chaotic Systems*, *Phys. Rev. A* **34**, 591 (1986).
- [44] R. Steinigeweg, J. Herbrych, and P. Prelovšek, *Eigenstate Thermalization within Isolated Spin-Chain Systems*, *Phys. Rev. E* **87**, 012118 (2013).
- [45] W. Beugeling, R. Moessner, and M. Haque, *Off-Diagonal Matrix Elements of Local Operators in Many-Body Quantum Systems*, *Phys. Rev. E* **91**, 012144 (2015).
- [46] D. Hetterich, M. Fuchs, and B. Trauzettel, *Equilibration in Closed Quantum Systems: Application to Spin Qubits*, *Phys. Rev. B* **92**, 155314 (2015).
- [47] A. Winter, *Coding Theorem and Strong Converse for Quantum Channels*, *IEEE Trans. Inf. Theory* **45**, 2481 (1999).
- [48] R. Bhatia, *Matrix Analysis* (Springer, New York, 1997), Vol. 169.
- [49] A. Riera, C. Gogolin, and J. Eisert, *Thermalization in Nature and on a Quantum Computer*, *Phys. Rev. Lett.* **108**, 080402 (2012).
- [50] A. J. Short, *Equilibration of Quantum Systems and Subsystems*, *New J. Phys.* **13**, 053009 (2011).